EXTREMAL PROPERTIES OF MARKOV CHAINS AND THE CONDITIONAL EXTREME VALUE MODEL

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

David Stephen Zeber

August 2012

EXTREMAL PROPERTIES OF MARKOV CHAINS AND THE CONDITIONAL EXTREME VALUE MODEL

David Stephen Zeber, Ph.D.

Cornell University 2012

Multivariate extreme value theory has proven useful for modeling multivariate data in fields such as finance and environmental science, where one is interested in accounting for the tendency of observations to exceed an extremely high (or low) threshold. Recent work has developed extremal models by studying the conditional distribution of a random vector, conditional on one of the components becoming extreme. This provides a way to handle situations such as asymptotic dependence, where traditional techniques may be uninformative. In this thesis, we explore the implications of the assumption that such a conditional distribution is well approximated by a limiting probability distribution when the conditioning component is extreme. We consider a version of the conditional distribution specified by a transition function.

If the transition kernel of a Markov chain satisfies our assumption, then a process known as the tail chain approximates the Markov chain over extreme states. We characterize the class of chains which admit such an approximation, and investigate the properties of the tail chain in relation to the distinction between extreme and non-extreme states. We find that, in general, the tail chain approximates a portion of the original process we term the "extremal component". We further derive the limit in distribution of a point process consisting of normalized Markov chain observations, expressing the limit in terms of the tail chain.

We also consider the case where a transition function satisfying our assumption describes the dependence structure of a random vector. We establish conditions under which a conditional extreme value model is appropriate, and derive the form of the limiting measure.

BIOGRAPHICAL SKETCH

David Zeber was born on September 20, 1984, in Ottawa, Ontario, Canada. After living for 10 years in Den Haag, Netherlands, he and his family moved back to Ottawa, where he graduated from Ashbury College in 2002. He received a Bachelor's degree in Mathematics (Honours) from McGill University in Montreal in 2006.

David joined the MS/PhD program in the Department of Statistical Science at Cornell University in 2006. Upon completing his PhD degree in July 2012, he joined Mozilla in San Francisco, California.

To my parents

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Professor Sidney Resnick. I am very grateful for his advice, guidance and encouragement through all stages of my graduate career. I learned a great deal about probability and statistics, and also about research, from a number of his courses that I had the opportunity to attend, as well as our weekly discussions. His many helpful comments and suggestions enhanced my thesis, papers, and talks. I very much appreciate his constant support, on both academic and personal levels, over the years I spent at Cornell.

I would also like to thank the other members of my special committee, Professors Gennady Samorodnitsky and Michael Nussbaum, for their support. Prof. Samorodnitsky's course in probability theory was a source of inspiration in my first year as a graduate student. Also, I learned a lot from Prof. Nussbaum's courses in mathematical statistics, and I am especially grateful to him for attending my B exam from Germany via videoconferencing.

I am grateful to have had the opportunity to pursue my graduate studies in the stimulating and supportive environments of both the Department of Statistical Science and the School of Operations Research and Information Engineering at Cornell. I would especially like to mention Professors Robert Strawderman and James Renegar, whose courses on statistical computing and mathematical programming I very much enjoyed, and Professors David Ruppert and Shane Henderson, from whom I learned a lot during my semesters as their teaching assistants. I would also like to thank Diana Drake and Beatrix Johnson for their assistance in a variety of matters.

The Department of Statistical Science has generously provided me with financial support during my six years at Cornell. I would like to thank Professor Robert

Strawderman for securing me funding during my final year. Additionally, my work with Professor Resnick was partially supported by ARO Contract W911NF-10-1-0289 and NSA Grant H98230-11-1-0193.

During my time as an undergraduate at McGill University, Professor David Wolfson taught me probability and statistics, and supervised my honours thesis. It is with his encouragement and support that I embarked upon a doctoral degree in statistics. I would also like to thank my high school math teacher, Glenna Allen, for her stimulating teaching, which inspired me to pursue a career in mathematics.

I am grateful to my friends for making my time in Ithaca enjoyable and memorable. Thanks to Raj, Inder and Michael for all the great memories and experiences, and for always being there for me. It was a pleasure to traverse each phase of my graduate career together with them. I would also like to thank Narisara for her constant support and encouragement.

Finally, I would like to thank my parents, Chris and Steve, for their support through the years.

TABLE OF CONTENTS

Biographical Sketch . . . iii
Dedication . . . iv
Acknowledgements . . . v
Table of Contents . . . vii
List of Symbols . . . ix

1 Introduction **1**
1.1 Extreme Values under Bivariate Dependence . . . 2
1.1.1 Markov Chain Models . . . 3
1.1.2 Conditional Extreme Value Models . . . 4
1.2 Technical Preliminaries . . . 4
1.2.1 Convergence of Measures . . . 4
1.2.2 Regular Variation . . . 6
1.2.3 Transition Functions and Markov Chains . . . 7

2 Asymptotics of Markov Kernels and the Tail Chain **9**
2.1 Overview . . . 9
2.2 Extremal Theory for Markov Kernels . . . 11
2.2.1 Tail Kernels . . . 12
2.2.2 Convergence to Tail Kernels . . . 13
2.2.3 Representation . . . 15
2.3 Convergence of Finite-Dimensional Distributions and the Extremal Component . . . 18
2.3.1 Distributions Conditional on the Initial State . . . 20
2.3.2 The Extremal Boundary . . . 22
2.3.3 The Extremal Component . . . 28
2.4 The Regularity Condition . . . 32
2.5 Examples . . . 39
2.6 Conclusions and Future Directions . . . 42

3 Convergence of Extremal Point Processes **45**
3.1 Overview . . . 45
3.2 Regular Variation Properties of the Extremal Component . . . 47
3.2.1 Effect of a Regularly Varying Initial Distribution . . . 47
3.2.2 Joint Tail Convergence . . . 50
3.2.3 Maximum of the Extremal Component . . . 56
3.3 Point Process Convergence for Markov Chains . . . 62
3.3.1 Cycle Decomposition . . . 62
3.3.2 General Point Process Convergence Results . . . 64
3.3.3 Cases where $G(\{0\}) = 0$ . . . 72
3.3.4 Cases where $G(\{0\}) > 0$ . . . 74
3.3.5 Jointly Regularly Varying Markov Chains . . . 77

3.4 Proof of Theorem 3.3.1 78
3.5 Conclusions and Future Directions 88

4 Conditional Extreme Value Modeling 91
4.1 Overview 91
4.2 Background 92
4.2.1 Technical Preliminaries: Extended Regular Variation 93
4.2.2 Technical Preliminaries: Domains of Attraction 94
4.2.3 The Conditional Extreme Value Model 95
4.3 Standard Case 97
4.3.1 Standard CEVM Properties 98
4.3.2 Examples 100
4.3.3 Counter-Examples 103
4.4 General Normalization for X 108
4.4.1 CEVM Properties 108
4.4.2 Standardization of X 113
4.4.3 Relation to the Heffernan and Tawn Model 119
4.5 General Normalizations for both X and Y 121
4.5.1 Kernel Asymptotics 122
4.5.2 CEVM Properties 126
4.6 Conclusions and Future Directions 129

A Technical Lemmas 132

Bibliography 137

LIST OF SYMBOLS

$\boldsymbol{x}_m$	The vector $(x_1, \ldots, x_m)$.
$f^{\leftarrow}$	The left-continuous inverse of a non-decreasing function f, i.e., $f^{\leftarrow}(x) = \inf\{y : f(y) \geq x\}$.
$f^{\rightarrow}$	The right-continuous inverse of a non-decreasing function f, i.e., $f^{\rightarrow}(x) = \inf\{y : f(y) > x\}$.
RV_ρ	The class of regularly varying functions with index ρ.
$\mathrm{ERV}_{\rho,k}$	The class of extended regularly varying functions with parameters ρ, k.
$\Pi(a)$	The class of Π-varying functions with auxiliary function a.
$D[0,\infty)$	The space of real-valued càdlàg functions on $[0,\infty)$ endowed with the Skorohod topology.
$D^{\uparrow}[0,\infty)$	The subspace of $D[0,\infty)$ consisting of non-decreasing functions f with $f(0) = 0$ and $\lim_{x\to\infty} f(x) = \infty$.
$\mathcal{B}(\mathbb{E})$	The Borel σ-field generated by open subsets of $\mathbb{E}$.
$\mathcal{K}(\mathbb{E})$	The collection of compact subsets of $\mathbb{E}$.
$\mathcal{C}(\mathbb{E})$	The space of real-valued continuous, bounded functions on $\mathbb{E}$.
$\mathcal{C}_K^+(\mathbb{E})$	The space of non-negative continuous functions on $\mathbb{E}$ with compact support.
$\mathbb{M}_+(\mathbb{E})$	The space of non-negative Radon measures on $\mathbb{E}$.
$\mathbb{M}_p(\mathbb{E})$	The space of Radon point measures on $\mathbb{E}$.
$\mathbb{LEB}$	Lebesgue measure.
$\mathrm{PRM}(\mu)$	Poisson random measure with mean measure μ.
ϵ_x	Point mass at x, i.e., $\epsilon_x(A) = \mathbf{1}_A(x)$.
ν_α	The measure on $(0,\infty]$ given by $\nu_\alpha(0,\infty] = x^{-\alpha}$ for $x > 0$, where $\alpha > 0$.
$\Rightarrow$	Weak convergence of probability measures.
$\xrightarrow{v}$	Vague convergence of Radon measures.
d_v	The vague metric, inducing the vague topology.
$\mu(f)$	The integral $\int f d\mu$.
$\mathbb{E}_m$	The cone $(0,\infty] \times [0,\infty]^m$.
$\mathbb{E}_m^*$	The cone $[0,\infty]^{m+1} \backslash \{\mathbf{0}\}$.
$\mathbb{E}_\gamma$	The interval $\{x \in \mathbb{R} : 1 + \gamma x > 0\}$.
$\overline{\mathbb{E}}_\gamma$	The closure on the right of the interval $\mathbb{E}_\gamma$.
$D(G_1^*)$	The standardized (Fréchet(1)) domain of attraction.

CHAPTER 1
INTRODUCTION

The theory of extreme values for independent, identically distributed samples is well-established and convenient. It is well-known that the asymptotic distribution of the maximum of n iid observations, appropriately normalized, belongs to a family parametrized by a single parameter, γ, under broad conditions on the distribution of the sample. These conditions translate to restrictions on the tail behaviour of the marginal distribution. In particular, both exponentially bounded tails (for which $\gamma = 0$), and tails which decay approximately as a power function (for which $\gamma > 0$), are included in this framework.

More recent work has explored extensions to this theory in cases where the observations are no longer independent. Two common settings are data consisting of multivariate observations, and observations from a stochastic process. In many cases, similar results still hold, although the extremal models become more complex, and possibly nonparametric. A central question is how the dependence structure affects extreme observations. Due to the dependence, extremely large or small observations may be more likely to occur together, compared to the iid case. For stochastic processes, a natural reformulation is the following: given that we have observed one extreme value (e.g., exceeding some high threshold), how many other extreme values are likely to occur in the "near future"?

Questions of this nature are readily phrased in terms of point processes. For example, the process $N_n(A) = \#\{j \in A : X_j > u_n\}$ counts the number of time-points in the set A at which the sequence $\boldsymbol{X} = \{X_j : j \geq 0\}$ exceeds the threshold u_n. As $n \to \infty$, such point processes typically converge in distribution to a limiting compound Poisson process. The compounding accounts for clustering of "nearby" extreme values, which occurs as a consequence of the dependence.

1.1 Extreme Values under Bivariate Dependence

We place ourselves in the framework of bivariate dependence, which we describe via a transition function K. This is a function that captures the stochastic behaviour of a pair of random variables (X, Y) by specifying, given an observed value $Y = y$, the subsequent probability distribution of X. Such a model is simple enough to admit tractable results, but is also useful in many practical settings. It is particularly suited to situations where X is known as an explicit (random) function of Y, such as in the regression problem $X = f(Y) + \varepsilon$.

In particular, we will consider cases where the distribution of X, conditional on $Y = y$, is well approximated by an asymptotic distribution under normalization, when the state y is extreme. Suppose $K : \mathbb{R} \times \mathcal{B}(\overline{\mathbb{R}}) \to [0, 1]$ is a transition function on the real line. Technical details are provided in Section 1.2.3; for now, it is sufficient to think of $K(y\,,\,A)$ as describing $\mathsf{P}[X \in A \,|\, Y = y]$. We work under the assumption that there exists a limiting probability distribution G on $\mathbb{R}$ such that

$$K\big(t\,,\,[-\infty, \alpha(t)x + \beta(t)]\big) \longrightarrow G[-\infty, x] \qquad \text{as } t \to \infty \tag{1.1.1}$$

at points of continuity of the limit $x \in \mathbb{R}$, for suitable normalizing functions $\alpha > 0$ and β. This means that the random fluctuation of X, when Y is extreme, is approximately described by the distribution G, up to normalization.

1.1.1 Markov Chain Models

In terms of stochastic processes, a formulation of conditional dependence in terms of transition functions applies directly to the study of Markov chains, whose dependence structure is completely specified by a transition kernel K. Markov chain models are widely used in applications to a variety of areas, including finance and environmental science. In particular, many commonly-used time series models can be expressed in terms of Markov chains on $\mathbb{R}^d$.

Previous authors have developed the "tail chain" as an extremal model for stationary Markov chains under the assumption (1.1.1). We investigate the tail chain model in Chapter 2. In general terms, our goals are to arrive at a deeper understanding of what conclusions may be gleaned from such an approximation, and to distinguish between the role of stationarity, which depends on the initial distribution, and the information inherent in the transition kernel, which determines the dependence structure. We introduce a precise notion of the distinction between extreme and non-extreme states, and we phrase the tail chain model in these terms. In addition, we obtain a characterization of the class of Markov chains for which this model is appropriate.

The tail chain model, if appropriate, captures all of the relevant information regarding dependence at extreme levels, meaning that it can be used to describe the clustering in a point process limit. In Chapter 3, we derive point process convergences, making explicit use of the regenerative structure inherent to a positive recurrent Markov chain. This departs from previous results, in which long-range dependence is generally controlled using mixing conditions.

1.1.2 Conditional Extreme Value Models

Aside from Markov chains, transition functions may be used to describe a functional dependence structure between components of a random vector, as mentioned above. We adopt this point of view in Chapter 4, where we consider the assumption (1.1.1) in the context of the Conditional Extreme Value Model (CEVM). The CEVM approximates the behaviour of a random vector when one component is extreme. We discuss conditions on the normalization functions α and β under which (1.1.1) leads to a CEVM, and we present formulas for the approximating limit measure in all cases. Moreover, we use the explicit dependence afforded by (1.1.1) to explore certain technical subtleties arising in the model.

1.2 Technical Preliminaries

We now review some technical concepts which will be used frequently in what follows.

1.2.1 Convergence of Measures

If not explicitly specified, we assume henceforth that any space $\mathbb{S}$ under discussion is a topological space paired with its Borel σ-field, $\mathcal{B}(\mathbb{S})$, generated by open sets. Denote by $\mathcal{K}(\mathbb{S})$ the collection of its compact sets; by $\mathcal{C}(\mathbb{S})$ the space of real-valued continuous, bounded functions on $\mathbb{S}$; and by $\mathcal{C}_K^+(\mathbb{S})$ the space of non-negative continuous functions with compact support.

If $\{P_n : n = 0, 1, \ldots\}$ are probability measures on a complete, separable space

$\mathbb{S}$, then P_n converges *weakly* to P_0, written $P_n \Rightarrow P_0$, if, as $n \to \infty$,

$$\int_{\mathbb{S}} f(x) P_n(dx) \longrightarrow \int_{\mathbb{S}} f(x) P_0(dx) \qquad \text{for all } f \in \mathcal{C}(\mathbb{S}). \tag{1.2.1}$$

Weak convergence can equivalently be phrased in terms of convergence on sets: $P_n \Rightarrow P_0$ iff $P_n(A) \Rightarrow P_0(A)$ for any $A \in \mathcal{B}(\mathbb{S})$ satisfying $P_0(\partial A) = 0$; this is an implication of the Portmanteau theorem [15, Theorem 2.1]. If $\{P_n\}$ are the respective distributions of a sequence of random variables $\{X_n\}$, (1.2.1) is equivalent to the convergence of the distribution functions $F_{X_n}(x) \to F_{X_0}(x)$ at points of continuity of the limit. For more information, see Billingsley [15].

For a space $\mathbb{E}$ which is locally compact with countable base (for example, a subset of $[-\infty, \infty]^d$), we denote by $\mathbb{M}_+(\mathbb{E})$ the space of non-negative Radon measures on $\mathbb{E}$. These are measures μ such that $\mu(K) < \infty$ for all $K \in \mathcal{K}(\mathbb{E})$. Given a sequence $\{\mu_n : n = 0, 1, \dots\}$ of measures in $\mathbb{M}_+(\mathbb{E})$, we say μ_n converges *vaguely* to μ_0, written $\mu_n \stackrel{v}{\to} \mu_0$, if, as $n \to \infty$,

$$\int_{\mathbb{E}} f(x) \mu_n(dx) \longrightarrow \int_{\mathbb{E}} f(x) \mu_0(dx) \qquad \text{for all } f \in \mathcal{C}_K^+(\mathbb{E}). \tag{1.2.2}$$

The convergence concept (1.2.2) induces a topology on $\mathbb{M}_+(\mathbb{E})$ which is metrizable by the vague metric, d_v, i.e., $d_v(\mu_n, \mu) \to 0$ iff $\mu_n \stackrel{v}{\to} \mu$. Refer to Resnick [76, Section 3.4] for further details.

A *random measure* is a random element of $\mathbb{M}_+(\mathbb{E})$. If $\{\eta_n : n = 0, 1, \dots\}$ are random measures, then $\eta_n \Rightarrow \eta_0$ in $\mathbb{M}_+(\mathbb{E})$ iff

$$\mathsf{E} \exp\{-\eta_n(f)\} \longrightarrow \mathsf{E} \exp\{-\eta_0(f)\} \qquad \text{for all } f \in \mathcal{C}_K^+(\mathbb{E}),$$

writing $\mu(f) = \int f d\mu$. Moreover, denote by ϵ_x the measure assigning unit mass to the point $x \in \mathbb{E}$, i.e., $\epsilon_x(A) = \mathbf{1}_A(x)$. A *point process* is a random element of $\mathbb{M}_p(\mathbb{E})$, the subspace of $\mathbb{M}_+(\mathbb{E})$ consisting of measures of the form $m = \sum_j \epsilon_{x_j}$. Note that $m(f) = \sum_j f(x_j)$. See Kallenberg [52] for more information.

Some useful technical results involving the convergence of integrals are assembled in the Appendix (p. 132).

1.2.2 Regular Variation

A measurable function $f : (0, \infty) \to (0, \infty)$ is *regularly varying* (at ∞) with index $\rho \in \mathbb{R}$, written $f \in \mathrm{RV}_\rho$, if

$$\lim_{t \to \infty} \frac{f(tx)}{f(t)} = x^\rho, \qquad x > 0. \tag{1.2.3}$$

Regular variation with $\rho = 0$ is called *slow variation*. If $\lim_{t\to\infty} f(tx)/f(t)$ exists for $x > 0$, then $f \in \mathrm{RV}_\rho$ for some ρ, and the limit is x^ρ. Also, if $f \in \mathrm{RV}_\rho$, then the convergence in (1.2.3) is locally uniform on $(0, \infty)$. In particular, this means that, given a function $x_t = x(t) \geq 0$ on $(0, \infty)$ such that $x(t) \to x > 0$ as $t \to \infty$, we have

$$\lim_{t \to \infty} \frac{f(tx_t)}{f(t)} = x^\rho$$

as well. Standard references include Bingham et al. [16] and Resnick [76].

Furthermore, given a random variable X on $[0, \infty)$ with distribution function F, we say that X has a *regularly varying tail* with *tail index* $\alpha > 0$ if $1 - F \in \mathrm{RV}_{-\alpha}$. Equivalently, there exists a scaling function $b(t) \to \infty$ such that $t\,\mathsf{P}[X > b(t)x] \to x^{-\alpha}$ for $x > 0$. This can be formulated equally in terms of vague convergence:

$$tF(b(t)\,\cdot) = t\,\mathsf{P}\left[\frac{X}{b(t)} \in \cdot\right] \xrightarrow{v} \nu_\alpha(\cdot) \quad \text{ in } \mathbb{M}_+(0, \infty] \quad \text{as } t \to \infty,$$

where ν_α is the measure given by $\nu_\alpha(x, \infty] = x^{-\alpha}$ for $x > 0$. Note that compact subsets of $(0, \infty]$ are contained in intervals of the form $[a, \infty]$. In higher dimensions, a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ has a *multivariate regularly varying tail* on a cone $\mathbb{E} \subset [-\infty, \infty]^d \backslash \{\mathbf{0}\}$ if there exists a non-degenerate Radon measure $\mu \in \mathbb{M}_+(\mathbb{E})$ and a scaling function $b(t) \to \infty$ such that

$$t\,\mathsf{P}\left[\frac{\boldsymbol{X}}{b(t)} \in \cdot\,\right] \xrightarrow{v} \mu(\cdot) \qquad \text{in } \mathbb{M}_+(\mathbb{E}) \quad \text{as } t \to \infty. \tag{1.2.4}$$

A stochastic process $\boldsymbol{X} = (X_0, X_1, \ldots)$ is *jointly regularly varying* if the finite-dimensional distributions $\mathsf{P}[(X_0, \ldots, X_m) \in \cdot\,]$ are regularly varying for each $m \geq 0$.

1.2.3 Transition Functions and Markov Chains

A *transition function* or *transition kernel* is a function $K : \mathbb{S} \times \mathcal{B}(\mathbb{S}') \to [0, 1]$ such that

(i) for each $A \in \mathcal{B}(\mathbb{S}')$, $K(\cdot\,, A)$ is a measurable function on $\mathbb{S}$; and

(ii) for each $y \in \mathbb{S}$, $K(y\,, \cdot)$ is a probability measure on $\mathbb{S}'$.

If $\mathbb{S} = \mathbb{S}'$, we say that K is a transition kernel on $\mathbb{S}$. Call K a *substochastic* transition function if $K(y\,, \mathbb{S}') \leq 1$ for each $y \in \mathbb{S}$. Also, note that for a random element (X, Y) of $\mathbb{S} \times \mathbb{S}'$, any particular version of the conditional distribution $\mathsf{P}[X \in \cdot \,|\, Y = y]$ is a transition function $K(y, \cdot)$.

A stochastic process $\boldsymbol{X} = (X_0, X_1, X_2, \ldots)$ on $\mathbb{S}$ is a (time-homogeneous) *Markov chain* with transition kernel $K : \mathbb{S} \times \mathcal{B}(\mathbb{S}) \to [0, 1]$, written $\boldsymbol{X} \sim K$, if its finite-dimensional distributions satisfy

$$\mathsf{P}\left[(X_{n+1}, \ldots, X_{n+m}) \in d\boldsymbol{x}_m \,\middle|\, X_n = x\right] = K(x\,, dx_1) \cdots K(x_{m-1}\,, dx_m), \quad n \geq 0,$$

for $m = 1, 2, \ldots$. We abbreviate $\boldsymbol{x}_m = (x_1, \ldots, x_m)$, similarly for $\boldsymbol{X}_m$, and we adopt the standard notation $\mathsf{P}_x[\boldsymbol{X}_m \in \cdot\,] = \mathsf{P}[\boldsymbol{X}_m \in \cdot \,|\, X_0 = x]$. Thus, if $\boldsymbol{X}$ has initial distribution $\mathsf{P}[X_0 \in \cdot\,] = F$, we have

$$\begin{aligned}\mathsf{P}_F\left[(X_0, \ldots, X_m) \in (dx_0, d\boldsymbol{x}_m)\right] &= F(dx_0)\,\mathsf{P}_{x_0}[\boldsymbol{X}_m \in d\boldsymbol{x}_m] \\ &= F(dx_0)K(x_0\,,\,dx_1)\cdots K(x_{m-1}\,,\,dx_m)\end{aligned}$$

for $m \geq 1$, and so $\mathsf{P}_x = \mathsf{P}_{\epsilon_x}$. The theory of Markov chains on general state spaces is given broad treatment by Meyn and Tweedie [62].

CHAPTER 2

ASYMPTOTICS OF MARKOV KERNELS AND THE TAIL CHAIN*

2.1 Overview

In studying the extremal properties of a stochastic process, it is useful to be able to describe the process's behaviour upon hitting an extreme state. Indeed, the dependence structure typically causes extremes to occur in clusters, raising questions such as how likely we are to observe other extreme values in the next few steps, and how many we are likely to observe. This information is conveniently summarized in cases where an approximation to the finite-dimensional distributions of the process over extreme states is appropriate, such as for multivariate regularly varying processes.

For a discrete-time Markov chain $\boldsymbol{X}$, the finite-dimensional distributions are approximated by an asymptotic process called the *tail chain*. This holds under an asymptotic assumption on the transition kernel of the chain. Loosely speaking, if the distribution of the next state converges under some normalization as the current state becomes extreme, then $\boldsymbol{X}$ behaves approximately as a multiplicative random walk, upon leaving a large initial state. This approach leads to intuitive extreme value models for processes such as autoregressions with random coefficients, which incorporate ARCH and other time series models. The focus on asymptotics for Markov kernels was introduced by Smith [87]. Perfekt [69, 70] ex-

*Material in this chapter has been accepted for publication in: S.I. Resnick and D. Zeber. Asymptotics of Markov kernels and the tail chain. *Advances in Applied Probability*, 45(1), 2013. Thanks to an anonymous referee for many helpful comments and suggestions on this material.

tended the approach to higher dimensions, and Segers [84] rephrased the conditions in terms of update functions. Basrak and Segers [8] extended the idea to a process approximation suitable for more general regularly varying stationary processes.

Markov chains which admit the tail chain approximation fall into one of two categories. Starting from an extreme state, the chain either remains extreme over any finite time horizon, or will drop to a "non-extreme" state of lower order after a finite amount of time. The latter case is problematic in that the tail chain model is not sensitive to possible subsequent jumps from a non-extreme state to an extreme one. This is usually handled by ruling out the class of processes exhibiting this behaviour via a technical condition, which we refer to as the *regularity condition*.

We find it more revealing to view the tail chain as approximating the restriction of the chain $\boldsymbol{X}$ to a certain portion of the state space. This is accomplished by formalizing the division between extreme and non-extreme states as a level we term the *extremal boundary*. We show that, in general, the tail chain approximates the *extremal component*, the portion of $\boldsymbol{X}$ having yet to cross below this boundary. Phrased in these terms, the regularity condition entails that the distinction between the original chain and its extremal component disappear asymptotically.

To this end, we propose an extreme value theory for Markov transition kernels, which leads directly to the convergence of finite-dimensional distributions conditional on the initial state, as it becomes extreme. In this context, it becomes clear that the regularity condition supplements certain missing information that is necessary to obtain a result for the complete chain $\boldsymbol{X}$. We phrase our theory, and relevant assumptions, equally in terms of update functions, as these often provide a more intuitive way of describing transition kernels, and are important in making connections with applications. In particular, most conditions are easily verified merely by inspecting the update function.

2.2 Extremal Theory for Markov Kernels

We begin by focusing on the Markov transition kernels rather than the stochastic processes they determine, and introduce a class of kernels we term "tail kernels," which we will view as scaling limits of certain kernels. See Section 1.2.3 (p. 7) for the definition of a transition kernel. Antecedents include Segers's definition of "back-and-forth tail chains" [84] that approximate certain Markov chains started from an extreme value.

For a Markov chain $\boldsymbol{X} \sim K$ on $[0, \infty)$, i.e., $\boldsymbol{X}$ transitions according to the transition kernel K (see Section 1.2.3), it is reasonable to expect that extremal behaviour of $\boldsymbol{X}$ is determined by pairs (X_n, X_{n+1}), and one way to control such pairs is to assume that (X_n, X_{n+1}) belongs to a bivariate domain of attraction. This approach was considered by Bortot and Coles [18] and Smith [87]. In the context of regular variation, writing

$$t\,\mathsf{P}\left[\frac{X_n}{b(t)} \in A_0\,,\ \frac{X_{n+1}}{b(t)} \in A_1\right] = \int_{A_0} K\big(b(t)u\,,\, b(t)A_1\big)\ t\,\mathsf{P}\left[\frac{X_n}{b(t)} \in du\right] \tag{2.2.1}$$

suggests combining marginal regular variation of X_n with a scaling kernel limit to derive extremal properties of the finite-dimensional distributions (fdds), as was done by Perfekt [69, 70] and Segers [84], and this is the direction we take. We first discuss the kernel scaling operation.

For simplicity, we assume the state space of the Markov chain is $[0, \infty)$, although with suitable modifications, it is relatively straightforward to extend the results to $\mathbb{R}^d$. Henceforth G will denote a general probability distribution on $[0, \infty)$.

2.2.1 Tail Kernels

First let us define the notion of tail kernel.

Definition. The *tail kernel associated with* G is given by

$$K^*(y\,,\,A) = \begin{cases} G(y^{-1}A) & y > 0 \\ \\ \epsilon_0(A) & y = 0 \end{cases} \tag{2.2.2}$$

for A measurable.

Recall ϵ_0 is the probability measure assigning unit mass to $\{0\}$. Thus, the class of tail kernels on $[0, \infty)$ is parameterized by probability distributions G. Such kernels are characterized by a scaling property:

Proposition 2.2.1. *A Markov transition kernel* K *is a tail kernel associated with some distribution* G *if and only if it satisfies the relation*

$$K(uy\,,\,A) = K(y\,,\,u^{-1}A) \tag{2.2.3}$$

with A *measurable and* $y \geq 0$ *for any* $u > 0$*, in which case* $G(\cdot) = K(1, \cdot)$.

Proof. Fix $u > 0$. If K is a tail kernel, (2.2.3) follows directly from the definition, using the fact that $\epsilon_0(A) = \epsilon_0(u^{-1}A)$ in the case $y = 0$. Conversely, assuming (2.2.3), for $y > 0$ we can write $K(y, A) = K(1, y^{-1}A)$, satisfying the first case of (2.2.2) with $G(\cdot) = K(1, \cdot)$. For the case $y = 0$, write $H(\cdot) = K(0, \cdot)$. We must show that $H(\cdot) = H(u^{-1}\cdot)$ implies $H = \epsilon_0$. Indeed, $H(0, \infty) = \lim_{n\to\infty} H(n^{-1}, \infty) = H(1, \infty)$, so $H(0, 1] = 0$. A similar argument shows that $H(1, \infty) = 0$ as well. $\square$

We call the Markov chain $\boldsymbol{T} \sim K^*$ the *tail chain associated with* G. Such a chain can be represented as

$$T_n = \xi_n \, T_{n-1} = T_0 \, \xi_1 \cdots \xi_n \qquad\qquad n = 1, 2, \ldots, \tag{2.2.4}$$

where $\xi_n \overset{\text{iid}}{\sim} G$ are independent of T_0. Thus, $\boldsymbol{T}$ is a multiplicative random walk with step distribution G and absorbing barrier at $\{0\}$.

2.2.2 Convergence to Tail Kernels

The tail chain approximates the behaviour of a Markov chain $\boldsymbol{X} \sim K$ in extreme states. Asymptotic results require that the normalized distribution of X_1 be well-approximated by some distribution G when X_0 is large, and we interpret this requirement as a domain of attraction condition for kernels.

Definition. A Markov transition kernel $K : [0, \infty) \times \mathcal{B}[0, \infty) \to [0, 1]$ is in the *domain of attraction of* G, written $K \in D(G)$, if as $t \to \infty$,

$$K(t \, , \, t \, \cdot) \Rightarrow G(\cdot) \qquad \text{on } [0, \infty]. \tag{2.2.5}$$

Note that $D(G)$ contains at least the tail kernel associated with G. A simple scaling argument extends (2.2.5) to

$$K(tu \, , \, t \, \cdot) \Rightarrow G(u^{-1} \cdot) = K^*(u \, , \, \cdot), \qquad u > 0, \tag{2.2.6}$$

where K^* is the tail kernel associated with G; this is the form appearing in (2.2.1). Thus tail kernels are scaling limits for kernels in a domain of attraction. In fact, tail kernels are the only possible limits:

Proposition 2.2.2. *Let K be a transition kernel on $[0, \infty)$. If there exists a family of probability distributions $\{G_u : u > 0\}$ on $[0, \infty)$ such that $K(tu \, , \, t \, \cdot) \Rightarrow G_u(\cdot)$ as*

$t \to \infty$for $u > 0$, then the function $\widehat{K}$ defined on $[0, \infty) \times \mathcal{B}[0, \infty)$ as

$$\widehat{K}(u\,,\, A) := \begin{cases} G_u(A) & u > 0 \\ \\ \epsilon_0(A) & u = 0 \end{cases}$$

is the tail kernel associated with G_1.

Proof. It suffices to show that $G_u(\cdot) = G_1(u^{-1}\cdot)$ for any $u > 0$. But this follows directly from the uniqueness of weak limits, since (2.2.6) shows that $K(tu, t\,\cdot) \Rightarrow G_1(u^{-1}\cdot)$. □

A version of (2.2.6) uniform in u is needed to derive convergence of finite-dimensional distributions.

Proposition 2.2.3. *Suppose $K \in D(G)$, and K^* is the tail kernel associated with G. Then, for any $u > 0$ and any non-negative function $u_t = u(t)$ such that $u_t \to u$ as $t \to \infty$, we have*

$$K(tu_t\,,\, t\,\cdot) \Rightarrow K^*(u\,,\,\cdot), \qquad (t \to \infty). \tag{2.2.7}$$

Proof. Suppose $u_t \to u > 0$. Observe that $K(tu_t, t\,\cdot) = K(tu_t, (tu_t)\, u_t^{-1}\cdot)$, and put $h_t(x) = u_t x$, $h(x) = ux$. Writing $P_t(\cdot) = K(tu_t, tu_t\,\cdot)$, we have

$$K(tu_t\,,\, t\,\cdot) = P_t \circ h_t^{-1} \Rightarrow G \circ h^{-1} = G(u^{-1}\cdot) = K^*(u\,,\,\cdot)$$

by [12, Theorem 5.5, p. 34]. □

The measure G controls $\boldsymbol{X}$ upon leaving an extreme state via (2.2.5). However, (2.2.5) is uninformative once $\boldsymbol{X}$ has reached a "non-extreme" state; indeed, (2.2.7) may fail if $u = 0$—see Example 2.5.2. The requirement $K^*(0, \cdot) = \epsilon_0$ reflects an assumption that a transition from non-extreme back to extreme is (asymptotically) impossible. The implications of such an assumption cannot be ignored if 0 is an accessible point of the state space, i.e., if $G(\{0\}) = K^*(y, \{0\}) > 0$.

One could conceivably extend this model to accomodate more general behaviour upon leaving a non-extreme state by allowing $K^*(0, A) = H(A)$ in (2.2.2), where H is a general probability distribution. The appropriate choice of H would then be determined via an asymptotic assumption analogous to (2.2.7), for example $K(tu_t, t\,\cdot) \Rightarrow H(\cdot)$ whenever $u_t \to 0$. In Section 2.4, we show that the "regularity condition" imposed by previous authors does in fact justify the choice $H = \epsilon_0$ in this manner (see (2.4.3), p. 33). We will not pursue this direction here. Instead, we proceed without assuming the regularity condition, viewing the prescription $H = \epsilon_0$ as "uninformative", in a sense to be made precise in Section 2.3.

Alternative formulations of (2.2.5) include replacing $K(t, t\,\cdot)$ with $K(t, a(t)\,\cdot)$ or else $K(t, a(t)\,\cdot + b(t))$, with appropriate functions $a(t) > 0$ and $b(t)$, in analogy with the usual domains of attraction conditions in extreme value theory, i.e.,

$$t\, \mathsf{P}\left[\frac{Y - b(t)}{a(t)} \in \cdot\right] \xrightarrow{v} \nu(\cdot) \qquad \text{in } \mathbb{M}_+(-\infty, \infty].$$

This direction will be explored in the context of the Conditional Extreme Value Model in Chapter 4. For now, we retain the standard choice $a(t) = t$, $b(t) = 0$.

2.2.3 Representation

How do we characterize kernels belonging to $D(G)$? From (2.2.4), for chains transitioning according to a tail kernel, the next state is a random multiple of the previous one, provided the prior state is non-zero. We expect that chains transitioning according to $K \in D(G)$ behave approximately like this upon leaving a large state, and this is best expressed in terms of a function describing how a new state depends on the prior one.

Given a kernel K, we can always find a sample space $\mathbb{E}$, a measurable function $\psi : [0, \infty) \times \mathbb{E} \to [0, \infty)$ and an $\mathbb{E}$-valued random element V such that

$$\psi(y, V) \stackrel{d}{=} K(y\,,\,\cdot), \qquad y \geq 0.$$

Given a random variable X_0, if we define the process $\boldsymbol{X} = (X_0, X_1, X_2, \dots)$ recursively as

$$X_{n+1} = \psi(X_n, V_{n+1}), \qquad n \geq 0,$$

where $\{V_n\}$ is an iid sequence equal in distribution to V and independent of X_0, then $\boldsymbol{X}$ is a Markov chain with transition kernel K. Call the function ψ an *update function corresponding to* K. If in addition $K \in D(G)$, the domain of attraction condition (2.2.5) becomes

$$t^{-1}\psi(t, V) \Rightarrow \xi,$$

where $\xi \sim G$. Applying the probability integral transform or the Skorohod representation theorems [13, Theorem 3.2, p. 6], [15, Theorem 6.7, p. 70] leads to the following result.

Proposition 2.2.4. *If K is a transition kernel on $[0, \infty)$, then $K \in D(G)$ if and only if there exists a measurable function $\psi^* : [0, \infty) \times [0, 1] \to [0, \infty)$ and a random variable $\xi^* \sim G$ on the uniform probability space $([0, 1], \mathcal{B}, \lambda)$ such that*

$$t^{-1}\psi^*(t, u) \longrightarrow \xi^*(u) \qquad \forall\, u \in [0, 1] \tag{2.2.8}$$

as $t \to \infty$, and ψ^* is an update function corresponding to K in the sense that

$$\lambda[\psi^*(y, \cdot) \in A] = K(y\,,\,A)$$

for measurable sets A.

Think of the update function as $\psi^*(y, U)$ where $U(u) = u$ is a uniform random variable on $[0, 1]$.

Proof. If there exist such ψ^* and ξ^* satisfying (2.2.8) then clearly $K \in D(G)$. Conversely, assume $K \in D(G)$, and let $\psi(\cdot, V)$ be an update function corresponding to K. According to Skorohod's representation theorem ([15, p. 70], with the necessary modifications to allow for an uncountable index set), there exists a random variable ξ^* and a stochastic process $\{Y_t^* ; t \geq 0\}$ defined on the uniform probability space $([0, 1], \mathcal{B}, \lambda)$, taking values in $[0, \infty)$, such that

$$\xi^* \sim G \,, \qquad Y_0^* \stackrel{d}{=} \psi(0, V) \,, \qquad Y_t^* \stackrel{d}{=} t^{-1}\psi(t, V) \;\text{ for } t > 0,$$

and $Y_t^*(u) \to \xi^*(u)$ as $t \to \infty$ for every $u \in [0, 1]$. Now, define $\psi^* : [0, \infty) \times [0, 1] \to [0, \infty)$ as

$$\psi^*(0, u) = Y_0^*(u) \qquad \text{and} \qquad \psi^*(t, u) = tY_t^*(u) \,, \quad t > 0, \qquad \forall\, u \in [0, 1].$$

It is evident that $\lambda[\psi^*(y, \cdot) \in A] = \mathsf{P}[\psi(y, V) \in A]$ for $y \in [0, \infty)$, so ψ^* is indeed an update function corresponding to K, and ψ^* satisfies (2.2.8) by construction. □

Update functions corresponding to K are not unique, and some of them may fail to converge pointwise as in (2.2.8). However (2.2.8) is convenient, and Proposition 2.2.4 shows that [84, Condition 2.2] in terms of update functions is equivalent to our weak convergence formulation $K \in D(G)$.

Pointwise convergence in (2.2.8) gives an intuitive representation of kernels in a domain of attraction.

Corollary 2.2.1. *$K \in D(G)$ iff there exists a random variable $\xi \sim G$ defined on the uniform probability space, and a measurable function $\phi : [0, \infty) \times [0, 1] \to$*

$(-\infty,\infty)$ *satisfying* $t^{-1}\phi(t,u) \to 0$ *for all* $u \in [0,1]$ *such that*

$$\psi(y,u) := \xi(u)\, y + \phi(y,u) \tag{2.2.9}$$

is an update function corresponding to K.

Proof. If such ξ and ϕ exist, then $t^{-1}\psi(t,u) = \xi(u) + t^{-1}\phi(t,u) \to \xi(u)$ for all u, so ψ satisfies (2.2.8). The converse follows from (2.2.8). □

Many Markov chains such as ARCH, GARCH and autoregressive processes are specified by structured recursions that allow quick recognition of update functions corresponding to kernels in a domain of attraction. A common example is the update function $\psi(y,(Z,W)) = Zy + W$, which behaves like $\psi'(y,Z) = Zy$ when y is large—compare ψ' to the form (2.2.4) discussed for tail kernels. In general, if K has an update function ψ of the form

$$\psi(y,(Z,W)) = Zy + \phi(y,W) \tag{2.2.10}$$

for a random variable $Z \geq 0$ and a random element W, where $t^{-1}\phi(t,w) \to 0$ whenever $w \in C$ for which $\mathsf{P}[W \in C] = 1$, then $K \in D(G)$ with $G = \mathsf{P}[Z \in \cdot\,]$. We will refer to update functions satisfying (2.2.10) as being in *canonical form*.

2.3 Convergence of Finite-Dimensional Distributions and the Extremal Component

Given a Markov chain $\boldsymbol{X} \sim K \in D(G)$, we show that the finite-dimensional distributions (fdds) of $\boldsymbol{X}$, started from an extreme state, converge to those of the tail chain $\boldsymbol{T}$ defined in (2.2.4).

We distinguish between two cases which represent substantially different types of behaviour. If $G(\{0\}) = 0$, observe that $\mathsf{P}[\boldsymbol{T}$ eventually hits $\{0\}] = 0$. On the other hand, if $G(\{0\}) > 0$, $\boldsymbol{T}$ hits $\{0\}$ in finite time with probability 1. In this case, the tail chain model is only appropriate up until the first hitting time of $\{0\}$. For example, consider the trajectory of $(X_1, \dots, X_m)$, started from $X_0 = t$, through the region $(t, \infty)^{m-2} \times [0, \delta] \times (t, \infty)$, where t is a high level. We would expect the tail chain to model this as a path through $(1, \infty)^{m-2} \times \{0\} \times (1, \infty)$. However, the probability of $\boldsymbol{T}$ taking paths through this set is 0, because $\{0\}$ is an absorbing state. Of course, if $G(\{0\}) = 0$, this restriction is moot.

It may be helpful to think instead in terms of the random walk $\{\log T_n\}$, for which the difference between the two cases amounts to whether or not the step distribution places positive mass at $\{-\infty\}$.

This raises the question of how to interpret the first hitting time of $\{0\}$ for $\boldsymbol{T}$ in terms of the original Markov chain $\boldsymbol{X}$. Such hitting times are important in the study of Markov chain point process models of exceedance clusters based on the tail chain, as discussed in Chapter 3. Intuitively, a transition to $\{0\}$ by $\boldsymbol{T}$ represents a transition from an extreme state to a non-extreme state by $\boldsymbol{X}$. We make this notion precise in Section 2.3.2 by viewing such transitions as downcrossings of a certain level we term the "extremal boundary."

Henceforth, $\boldsymbol{X}$ is a Markov chain on $[0, \infty)$ with transition kernel $K \in D(G)$, K^* is the tail kernel associated with G, and $\boldsymbol{T}$ is a Markov chain on $[0, \infty)$ with kernel K^* (see Section 1.2.3, p. 7, for further details). Recall the shorthand $\boldsymbol{x}_m = (x_1, \dots, x_m)$, and similarly for $\boldsymbol{X}_m$ and $\boldsymbol{T}_m$.

2.3.1 Distributions Conditional on the Initial State

Define for $m \geq 1$ the conditional distributions

$$\pi_m^{(t)}(u\,,\,\cdot) = \mathsf{P}_{tu}\left[\left(\frac{X_1}{t}, \ldots, \frac{X_m}{t}\right) \in \cdot\,\right] \quad \text{and} \quad \pi_m(u\,,\,\cdot) = \mathsf{P}_u\left[(T_1, \ldots, T_m) \in \cdot\,\right] \tag{2.3.1}$$

on $[0,\infty) \times \mathcal{B}[0,\infty]^m$. We consider when $\pi_m^{(t)} \Rightarrow \pi_m$ on $[0,\infty]^m$ pointwise in u, i.e., when the domain of attraction condition (2.2.5) extends to the finite-dimensional conditional distributions. If $G(\{0\}) = 0$, this is a direct consequence of the domain of attraction condition (2.2.5), but if $G(\{0\}) > 0$, more thought is required. We begin by restricting the convergence to the smaller space $\mathbb{E}'_m := (0,\infty]^{m-1} \times [0,\infty]$. Relatively compact sets in $\mathbb{E}'_m$ are contained in rectangles $[\boldsymbol{a}, \boldsymbol{\infty}] \times [0,\infty]$, where $\boldsymbol{a} \in (0,\infty)^{m-1}$.

Theorem 2.3.1. *Suppose $\boldsymbol{X} \sim K$ and $\boldsymbol{T} \sim K^*$ are Markov chains on $[0,\infty)$, where $K \in D(G)$ and K^* is the tail kernel associated with G, and recall the conditional distributions $\pi_m^{(t)}$ and π_m defined in (2.3.1). Let $u_t = u(t)$ be a non-negative function such that $u_t \to u > 0$ as $t \to \infty$.*

(a) The restrictions to $\mathbb{E}'_m$,

$$\mu_m^{(t)}(u\,,\,\cdot\,) := \pi_m^{(t)}(u\,,\,\cdot \cap \mathbb{E}'_m) \quad \textit{and} \quad \mu_m(u\,,\,\cdot\,) := \pi_m(u\,,\,\cdot \cap \mathbb{E}'_m), \tag{2.3.2}$$

satisfy

$$\mu_m^{(t)}(u_t\,,\,\cdot\,) \xrightarrow{v} \mu_m(u\,,\,\cdot\,) \qquad \textit{in } \mathbb{M}_+(\mathbb{E}'_m) \qquad (t \to \infty). \tag{2.3.3}$$

(b) If $G(\{0\}) = 0$, we have

$$\pi_m^{(t)}(u_t\,,\,\cdot\,) \Rightarrow \pi_m(u\,,\,\cdot\,) \qquad \textit{on } [0,\infty]^m \qquad (t \to \infty). \tag{2.3.4}$$

Proof. Exploiting the Markov structure, we proceed via an induction argument facilitated by Lemma A.0.2 (p. 132). Consider (a) first. If $m = 1$, then (2.3.3) above reduces to (2.2.7). Assume $m \geq 2$, and let $f \in \mathcal{C}_K^+(\mathbb{E}'_m)$. Writing $\mathbb{E}'_m = (0,\infty] \times \mathbb{E}'_{m-1}$, we can find $a > 0$ and $B \in \mathcal{K}(\mathbb{E}'_{m-1})$ such that f is supported on $[a,\infty] \times B$. Now, observe that

$$\begin{aligned}\mu_m^{(t)}(u_t\,,\,\cdot)(f) &= \int_{(0,\infty]} K(tu_t\,,\,tdx_1) \int_{\mathbb{E}'_{m-1}} K(tx_1\,,\,tdx_2)\cdots K(tx_{m-1}\,,\,tdx_m)\; f(\boldsymbol{x}_m) \\ &= \int_{(0,\infty]} K(tu_t\,,\,tdx_1) \int_{\mathbb{E}'_{m-1}} \mu_{m-1}^{(t)}(x_1\,,\,d(x_2,\ldots,x_m))\; f(\boldsymbol{x}_m).\end{aligned}$$

Defining

$$h_t(v) = \int_{\mathbb{E}'_{m-1}} \mu_{m-1}^{(t)}(v\,,\,d\boldsymbol{x}_{m-1})\; f(v,\boldsymbol{x}_{m-1})$$

and

$$h(v) = \int_{\mathbb{E}'_{m-1}} \mu_{m-1}(v\,,\,d\boldsymbol{x}_{m-1})\; f(v,\boldsymbol{x}_{m-1})\,,$$

the previous expression becomes

$$\mu_m^{(t)}(u_t\,,\,\cdot)(f) = \int_{(0,\infty]} K(tu_t\,,\,tdv)\; h_t(v).$$

Now, suppose $v_t \to v > 0$; we verify

$$h_t(v_t) \longrightarrow h(v). \tag{2.3.5}$$

By continuity, we have $f(v_t, \boldsymbol{x}_{m-1}^t) \to f(v, \boldsymbol{x}_{m-1})$ whenever $\boldsymbol{x}_{m-1}^t \to \boldsymbol{x}_{m-1}$, and the induction hypothesis provides $\mu_{m-1}^{(t)}(v_t,\cdot) \xrightarrow{v} \mu_{m-1}(v,\cdot)$. Also, $f(x,\cdot)$ has compact support B (without loss of generality, $\mu_{m-1}(v,\partial B) = 0$). Combining these facts, (2.3.5) follows from Lemma A.0.2 (b). Next, since the h_t and h have common compact support $[a,\infty]$, and recalling from Proposition 2.2.3 that $K(tu_t, t\,\cdot) \Rightarrow K^*(u,\cdot)$, Lemma A.0.2 (a) yields

$$\mu_m^{(t)}(u_t\,,\,\cdot)(f) \longrightarrow \int_{(0,\infty]} K^*(u\,,\,dv)\; h(v) = \mu_m(u\,,\,\cdot)(f).$$

Implication (b) follows from essentially the same argument. For $m \geq 2$, suppose $f \in \mathcal{C}[0,\infty]^m$. Replacing μ by π and $\mathbb{E}'_{m-1}$ by $[0,\infty]^{m-1}$ in the definitions of h_t and h, we have

$$\pi_m^{(t)}(u_t\,,\,\cdot)(f) = \int_{[0,\infty]} K(tu_t\,,\,tdv)\; h_t(v).$$

This time Lemma A.0.2 (a) shows that $h_t(v_t) \to h(v)$ if $v_t \to v > 0$, and since $K^*(u,(0,\infty]) = 1$, resorting to Lemma A.0.2 (a) once more yields

$$\pi_m^{(t)}(u_t\,,\,\cdot)(f) \longrightarrow \int_{[0,\infty]} K^*(u\,,\,dv)\; h(v) = \pi_m(u\,,\,\cdot)(f). \qquad \square$$

If $G(\{0\}) > 0$, then $K^*(u,(0,\infty]) = 1 - G(\{0\}) < 1$, and for (2.3.4) to hold would require knowing the behaviour of $h_t(v_t)$ when $v_t \to 0$ as well. Previous work handled this using the regularity condition discussed in Section 2.4 (p. 32).

2.3.2 The Extremal Boundary

The normalization employed in the domain of attraction condition (2.2.5) suggests that, starting from a large state t, the extreme states are on the order of scalar multiples of t. For example, we would consider a transition from t into $(t/3, 2t]$ to remain extreme. Thus, we think of states which can be made smaller than $t\delta$ for any δ, if t is large enough, as non-extreme. In this context, the set $[0, \sqrt{t}]$ would consist of non-extreme states.

Asymptotically, a tail chain path through $(0,\infty)$ models the original chain $\boldsymbol{X}$ as it travels among extreme states, and all of the non-extreme states are compacted into the state $\{0\}$ in the state space of $\boldsymbol{T}$. Since $\{0\}$ is an absorbing barrier for $\boldsymbol{T}$, the tail chain is informative as a model only up until the first time $\boldsymbol{X}$ crosses to a non-extreme state. Since a transition of $\boldsymbol{X}$ from extreme to non-extreme is very unlikely if $G(\{0\}) = 0$, the tail chain captures all the relevant extremal behaviour of $\boldsymbol{X}$ (Theorem 2.3.1 (b)).

Drawing upon this interpretation, we develop a rigorous formulation of the distinction between extreme and non-extreme states, and we recast Theorem 2.3.1 as convergence on the unrestricted space $[0, \infty]^m$ of the conditional distributions, given that $\boldsymbol{X}$ has not yet reached a non-extreme state.

Definition. Suppose $K \in D(G)$. An *extremal boundary* for K is a non-negative function $y(t)$ defined on $[0, \infty)$, satisfying $\lim_{t\to\infty} y(t) = 0$ and

$$K(t\,,\, t\,[0, y(t)]) \longrightarrow G(\{0\}) \qquad \text{as } \ t \to \infty. \tag{2.3.6}$$

Such a function is guaranteed to exist by Lemma A.0.5 (p. 135).

If $G(\{0\}) = 0$, then $y(t) \equiv 0$ is a trivial choice. For any function $0 \le y(t) \to 0$, we have $\limsup_{t\to\infty} K(t, t\,[0, y(t)]) \le G(\{0\})$, so (2.3.6) is equivalent to

$$\liminf_{t\to\infty} K(t\,,\, t\,[0, y(t)]) \ge G(\{0\}). \tag{2.3.7}$$

If $y(t)$ is an extremal boundary, it follows that any function $0 \le \tilde{y}(t) \to 0$ with $\tilde{y}(t) \ge y(t)$ for $t \ge t_0$ is also an extremal boundary for K. Taking $\tilde{y}(t) = \vee_{s\ge t}\, y(s)$ shows that without loss of generality, we can assume $y(t)$ to be non-increasing.

The extremal boundary has a natural formulation in terms of the update function. As in (2.2.10), let $\psi(y, (Z, W)) = Zy + \phi(y, W)$ be an update function in canonical form, where y is extreme. If $Z > 0$ then the next state is approximately Zy, another extreme state. Otherwise, if $Z = 0$, the next state is $\phi(y, W)$, and a transition from an extreme to a non-extreme state has taken place. This suggests choosing an extremal boundary whose order is between t and $\phi(t, w)$.

Proposition 2.3.1. *Suppose* $\psi(y,(Z,W))$ *is an update function in canonical form as in* (2.2.10)*. If* $\zeta(t) > 0$ *is a function on* $[0,\infty)$ *such that*

$$\frac{\phi(t,w)}{\zeta(t)} \longrightarrow 0 \tag{2.3.8}$$

as $t \to \infty$ *whenever* $w \in B$ *for which* $\mathsf{P}[W \in B] = 1$*, then*

$$\liminf_{t\to\infty} K(t,[0,\zeta(t)]) \geq G(\{0\}).$$

Provided $\lim_{t\to\infty} \zeta(t)/t = 0$*, an extremal boundary is given by* $y(t) := \zeta(t)/t$*.*

Thus if $\phi(t,w) = o(\zeta(t))$ and $\zeta(t) = o(t)$ then $\zeta(t)/t$ is an extremal boundary. For example, if $\psi(y,(Z,W)) = Zy + W$, so that $\phi(t,w) = w$, then choosing $\zeta(t)$ to be any function $\zeta(t) \to \infty$ such that $\zeta(t) = o(t)$ makes $\zeta(t)/t$ an extremal boundary. Taking $\zeta(t) = \sqrt{t}$, we find that $y(t) = 1/\sqrt{t}$ is an extremal boundary.

Proof. Since

$$\begin{aligned}
\mathsf{P}[\psi(t,(Z,W)) \leq \zeta(t),\ Z = 0] &= \mathsf{P}[\phi(t,W) \leq \zeta(t),\ Z = 0] \\
&\geq \mathsf{P}[|\phi(t,W)| \leq \zeta(t),\ Z = 0] \\
&\geq \mathsf{P}[Z = 0] - \mathsf{P}\left[\frac{|\phi(t,W)|}{\zeta(t)} > 1\right] \longrightarrow \mathsf{P}[Z = 0],
\end{aligned}$$

we have

$$\liminf_{t\to\infty} K(t\,,\ [0,\zeta(t)]) = \liminf_{t\to\infty} \mathsf{P}[\psi(t,(Z,W)) \leq \zeta(t)] \geq \mathsf{P}[Z = 0]. \qquad \square$$

We will need an extremal boundary for which (2.3.6) still holds upon replacing the initial state t with tu_t, where $u_t \to u > 0$. Compare the following extension with Proposition 2.2.3.

Proposition 2.3.2. *If $K \in D(G)$, then there exists an extremal boundary $y^*(t)$ such that*

$$K(tu_t\,,\, t\,[0, y^*(t)]) \longrightarrow G(\{0\}) \qquad \text{as } \; t \to \infty \tag{2.3.9}$$

for any non-negative function $u_t = u(t) \to u > 0$.

We will refer to y^* as a *uniform extremal boundary.*

Proof. Let $y(t)$ be an extremal boundary for K. As a first step, fix $u_0 > 1$, and suppose $u_0^{-1} < u < u_0$. Define $\tilde{y}(t) = u_0\, y(tu_0^{-1})$. Now, if $u_t \to u$, then $y_{\{u\}}(t) := u_t\, y(tu_t)$ satisfies (2.3.9), since

$$K\big(tu_t\,,\, t\,[0, y_{\{u\}}(t)]\big) = K\big(tu_t\,,\, tu_t\,[0, y(tu_t)]\big) \longrightarrow G(\{0\}).$$

Here $y_{\{u\}}$ depends on the choice of function u_t. However, since we eventually have $u_0^{-1} < u_t < u_0$ for t large enough, it follows that $\tilde{y}(t) > y_{\{u\}}(t)$ for such t. Hence, $\tilde{y}(t)$ satisfies (2.3.9) for any $u_t \to u$ with $u_0^{-1} < u < u_0$.

We now remove the restriction in u_0 via a diagonalization argument. For $k = 2, 3, \ldots$, let $y_k(t)$ be extremal boundaries such that $K(tu_t, t\,[0, y_k(t)]) \to G(\{0\})$ whenever $u_t \to u$ for $u \in (k^{-1}, k)$, and put $y_0 = y_1 = y$. Next, define the sequence $\{(s_k, x_k) : k = 0, 1, \ldots\}$ inductively as follows. Setting $s_0 = 0$ and $x_0 = y_0(1)$, choose $s_k \geq s_{k-1} + 1$ such that $y_j(t) \leq k^{-1} \wedge x_{k-1}$ for all $j = 0, \ldots, k$ whenever $t \geq s_k$, and put $x_k = \max\{y_j(s_k) : j = 0, \ldots, k\}$. Note that $x_k \leq k^{-1} \wedge x_{k-1}$, so $x_k \downarrow 0$, and $s_k \uparrow \infty$. Finally, set

$$y^*(t) = \sum_{k=0}^{\infty} x_k\, \mathbf{1}_{[s_k,\, s_{k+1})}(t).$$

Observe that $0 \leq y^*(t) \downarrow 0$, and suppose $u_t \to u > 0$. Then $u \in (k_0^{-1}, k_0)$ for some k_0, so $K(tu_t, t\,[0, y_{k_0}(t)]) \to G(\{0\})$, and for $k \geq k_0$, our construction ensures that whenever $s_k \le t < s_{k+1}$, we have $y_{k_0}(t) \le y_{k_0}(s_k) \le x_k = y^*(t)$. Therefore, $y^*(t) \ge y_{k_0}(t)$ for $t \ge s_{k_0}$, so y^* satisfies (2.3.9). □

Henceforth, we assume any $K \in D(G)$ is accompanied by a uniform extremal boundary denoted by $y(t)$, and we consider extreme states on the order of t to be $(ty(t), \infty]$. If $G(\{0\}) = 0$, then all positive states are extreme states. We now use the extremal boundary to reformulate the convergence of Theorem 2.3.1 on the larger space $[0,\infty]^m$. Put $\mathbb{E}'_m(t) = (y(t),\infty]^{m-1} \times [0,\infty]$, so that $\mathbb{E}'_m(t) \uparrow \mathbb{E}'_m = (0,\infty]^{m-1} \times [0,\infty]$. Recall the notation $\mu_m^{(t)}$ and μ_m^* from (2.3.1), (2.3.2) in Theorem 2.3.1 (p. 20).

Theorem 2.3.2. *Let $u_t = u(t)$ be a non-negative function such that $u_t \to u > 0$ as $t \to \infty$. Taking*

$$\widetilde{\mu}_m^{(t)}(u\,,\,\cdot\,) = \pi_m^{(t)}(u\,,\,\cdot\,\cap \mathbb{E}'_m(t)),$$

we have

$$\widetilde{\mu}_m^{(t)}(u_t\,,\,\cdot\,) \xrightarrow{v} \mu_m(u\,,\,\cdot\,) \qquad \text{in } \mathbb{M}_+[0,\infty]^m \qquad (t \to \infty).$$

Proof. Note that we can just as well write $\widetilde{\mu}_m^{(t)}(u,\cdot) = \mu_m^{(t)}(u, \cdot \cap \mathbb{E}'_m(t))$. Suppose $m \ge 2$ and let $f \in \mathcal{C}_K^+[0,\infty]^m$. For $\delta > 0$, define $A_\delta = (\delta,\infty]^{m-1} \times [0,\infty]$, and choose δ such that $\mu_m(u, \partial A_\delta) = 0$. On the one hand, for large t we have

$$\begin{aligned}\widetilde{\mu}_m^{(t)}(u_t\,,\,\cdot)(f) &= \int_{[0,\infty]^m} f(\boldsymbol{x})\,\mathbf{1}_{\mathbb{E}'_m(t)}(\boldsymbol{x})\,\mu_m^{(t)}(u_t\,,\,d\boldsymbol{x}) \\ &\ge \int_{\mathbb{E}'_m} f(\boldsymbol{x})\,\mathbf{1}_{A_\delta}(\boldsymbol{x})\,\mu_m^{(t)}(u_t\,,\,d\boldsymbol{x}) \longrightarrow \int_{\mathbb{E}'_m} f(\boldsymbol{x})\,\mathbf{1}_{A_\delta}(\boldsymbol{x})\,\mu_m(u\,,\,d\boldsymbol{x})\end{aligned}$$

as $t \to \infty$ by Lemma A.0.3 (p. 133). Letting $\delta \downarrow 0$ yields

$$\liminf_{t\to\infty} \widetilde{\mu}_m^{(t)}(u_t\,,\,\cdot)(f) \;\ge\; \mu_m(u\,,\,\cdot)(f) \tag{2.3.10}$$

by monotone convergence. On the other hand, fixing δ, we can decompose the space according to the first downcrossing of δ:

$$\begin{aligned} &\widetilde{\mu}_m^{(t)}(u_t\,,\,\cdot)(f) \qquad\qquad (2.3.11)\\ &\quad= \int_{[0,\infty]^m} f(\boldsymbol{x})\,\mathbf{1}_{A_\delta}(\boldsymbol{x})\,\widetilde{\mu}_m^{(t)}(u_t\,,\,d\boldsymbol{x}) + \sum_{k=1}^{m-1}\int_{[0,\infty]^m} f(\boldsymbol{x})\,\mathbf{1}_{A_\delta^k}(\boldsymbol{x})\,\widetilde{\mu}_m^{(t)}(u_t\,,\,d\boldsymbol{x}), \end{aligned}$$

where $A_\delta^k = (\delta,\infty]^{k-1}\times[0,\delta]\times[0,\infty]^{m-k}$. On the subsets A_δ^k we appeal to the bound on f, say M, to obtain

$$\int_{[0,\infty]^m} f(\boldsymbol{x})\,\mathbf{1}_{A_\delta^k}(\boldsymbol{x})\,\widetilde{\mu}_m^{(t)}(u_t\,,\,d\boldsymbol{x}) \le M\widetilde{\mu}_m^{(t)}(u_t\,,\,A_\delta^k).$$

Now,

$$\begin{aligned} \widetilde{\mu}_m^{(t)}(u_t\,,\,A_\delta^k) &\le \mu_k^{(t)}(u_t\,,\,(\delta,\infty]^{k-1}\times(y(t),\delta]) \qquad\qquad (2.3.12)\\ &= \mu_k^{(t)}(u_t\,,\,(\delta,\infty]^{k-1}\times[0,\delta]) - \mu_k^{(t)}(u_t\,,\,(\delta,\infty]^{k-1}\times[0,y(t)]). \end{aligned}$$

Considering the second term, we have

$$\begin{aligned} &\mu_k^{(t)}(u_t\,,\,(\delta,\infty]^{k-1}\times[0,y(t)])\\ &\quad= \int_{[0,\infty]} K(tu_t\,,\,tdx_1)\,\mathbf{1}_{(\delta,\infty]}(x_1)\cdots\int_{[0,\infty]} K(tx_{k-2}\,,\,tdx_{k-1})\,\mathbf{1}_{(\delta,\infty]}(x_{k-1})\\ &\qquad\qquad\qquad\qquad\qquad\qquad\cdot K(tx_{k-1}\,,\,t\,[0,y(t)])\\ &\quad= \int_{\mathbb{E}'_{k-1}} \mu_{k-1}^{(t)}(u_t\,,\,d\boldsymbol{x}_{k-1})\;h_t(\boldsymbol{x}_{k-1}), \end{aligned}$$

where

$$h_t(\boldsymbol{x}_{k-1}) = K(tx_{k-1}\,,\,t\,[0,y(t)])\,\mathbf{1}_{(\delta,\infty]^{k-1}}(\boldsymbol{x}_{k-1}).$$

Moreover, if $\boldsymbol{x}_{k-1}^t \to \boldsymbol{x}_{k-1} \in (\delta,\infty]^{k-1}$, then

$$h_t(\boldsymbol{x}_{k-1}^t) = K(tx_{k-1}^t\,,\,t\,[0,y(t)])\,\mathbf{1}_{(\delta,\infty]^{k-1}}(\boldsymbol{x}_{k-1}^t) \longrightarrow G(\{0\})\,\mathbf{1}_{(\delta,\infty]^{k-1}}(\boldsymbol{x}_{k-1}),$$

using the fact that $y(t)$ is a uniform extremal boundary. Since

$$\mu_{k-1}(u,\partial(\delta,\infty]^{k-1}) = 0$$

without loss of generality by choice of δ, we conclude that

$$\mu_k^{(t)}(u_t\,,\,(\delta,\infty]^{k-1}\times[0,y(t)])\longrightarrow G(\{0\})\cdot\mu_{k-1}(u\,,\,(\delta,\infty]^{k-1})$$
$$=\mu_k(u\,,\,(\delta,\infty]^{k-1}\times\{0\})$$

as $t\to\infty$. Now, let us return to (2.3.12). Given any $\epsilon>0$, by choosing δ small enough, we can make

$$\begin{aligned}
&\mu_k^{(t)}(u_t\,,\,(\delta,\infty]^{k-1}\times(y(t),\delta])\\
&\quad\longrightarrow\mu_k(u\,,\,(\delta,\infty]^{k-1}\times[0,\delta])-\mu_k(u\,,\,(\delta,\infty]^{k-1}\times\{0\})\\
&\quad\le\mu_k(u\,,\,(0,\infty]^{k-1}\times[0,\delta])-\mu_k(u\,,\,(\delta,\infty]^{k-1}\times\{0\})\\
&\quad<\mu_k(u\,,\,(0,\infty]^{k-1}\times\{0\})+\frac{\epsilon}{2}-\left\{\mu_k(u\,,\,(0,\infty]^{k-1}\times\{0\})-\frac{\epsilon}{2}\right\}=\epsilon,
\end{aligned}$$

i.e.,

$$\limsup_{t\to\infty}\widetilde{\mu}_m^{(t)}(u_t\,,\,A_\delta^k)<\epsilon,\qquad k=1,\dots,m-1. \tag{2.3.13}$$

Therefore, (2.3.11) implies that, given $\epsilon'>0$,

$$\begin{aligned}
\limsup_{t\to\infty}\widetilde{\mu}_m^{(t)}(u_t\,,\,\cdot)(f)\le\int_{[0,\infty]^m}f(\boldsymbol{x})\,\mathbf{1}_{A_\delta}(\boldsymbol{x})\;\mu_m(u\,,\,d\boldsymbol{x})&\\
+\,M\sum_{k=1}^{m-1}\limsup_{t\to\infty}\widetilde{\mu}_m^{(t)}(u_t\,,\,A_\delta^k)&\\
<\mu_m(u\,,\,\cdot)(f)+\epsilon'&
\end{aligned}$$

for small enough δ. Combining this with (2.3.10) yields the result. □

2.3.3 The Extremal Component

Having thus formalized the distinction between extreme and non-extreme states, we return to the question of phrasing a general fdd limit result for $\boldsymbol{X}$. The extremal boundary allows us to interpret the first hitting time of $\{0\}$ by the tail chain as approximating the time of the first transition from extreme down to non-extreme. In this terminology, Theorem 2.3.2 provides a result, given that such a transition has yet to occur.

Define the first hitting time of a non-extreme state

$$\tau(t) = \inf\{n \geq 0 : X_n \leq ty(t)\}\,.$$

For a Markov chain started from tu_t, where $u_t \to u > 0$, we have $tu_t > ty(t)$ for large t, so $\tau(t)$ is the first downcrossing of the extremal boundary.

For the tail chain $\boldsymbol{T} = \{T_0\xi_1 \cdots \xi_n : n = 0, 1, \dots\}$, put

$$\tau^* = \inf\{n \geq 0 : T_n = 0\}.$$

Given $T_0 > 0$, $\tau^* = \inf\{n \geq 1 : \xi_n = 0\}$, i.e., τ^* follows a Geometric distribution with parameter $p = G(\{0\})$. Thus, $\mathsf{P}[\tau^* = m] = p(1-p)^{m-1}$ for $m \geq 1$ if $p > 0$, and $\mathsf{P}[\tau^* = \infty] = 1$ if $p = 0$. Theorem 2.3.2 becomes

$$\mathsf{P}_{tu_t}\left[\frac{\boldsymbol{X}_m}{t} \in \cdot\,,\ \tau(t) \geq m\right] \xrightarrow{v} \mathsf{P}_u[\boldsymbol{T}_m \in \cdot\,,\ \tau^* \geq m], \tag{2.3.14}$$

implying that τ^* approximates $\tau(t)$:

$$\mathsf{P}_{tu_t}[\tau(t) \in \cdot\,] \Rightarrow \mathsf{P}_u[\tau^* \in \cdot\,], \qquad (t \to \infty,\ u_t \to u > 0). \tag{2.3.15}$$

So if $G(\{0\}) > 0$, $\boldsymbol{X}$ takes an average of approximately $G(\{0\})^{-1}$ steps to return to a non-extreme state. However, if $G(\{0\}) = 0$, $\mathsf{P}_{tu_t}[\tau_1 \leq m] \to 0$ for any $m \geq 1$; in other words, starting from a larger and larger initial state, it will take longer and longer for $\boldsymbol{X}$ to cross down to a non-extreme state.

We now restate (2.3.14) in terms of a process derived from $\boldsymbol{X}$, called the *extremal component* of $\boldsymbol{X}$, whose unrestricted fdds converge weakly to those of $\boldsymbol{T}$.

Definition. The *extremal component* of $\boldsymbol{X}$ relative to t is the process $\boldsymbol{X}^{(t)}$ defined for $t > 0$ as

$$X_n^{(t)} = X_n \cdot \mathbf{1}_{\{n<\tau(t)\}}\,, \qquad n = 0, 1, \ldots.$$

Thus, the extremal component $\boldsymbol{X}^{(t)}$ consists of the portion of the process $\boldsymbol{X}$ up until the first transition to a non-extreme state. Note that $\tau(t)$ can be expressed as

$$\tau(t) = \inf\{n \geq 0 : X_n^{(t)} = 0\}.$$

Observe that $\boldsymbol{X}^{(t)}$ is a Markov chain on $[0, \infty)$ with transition kernel

$$K^{(t)}(x\,,\,A) = \begin{cases} K\big(x\,,\,A \cap (ty(t), \infty]\big) + \epsilon_0(A) \cdot K\big(x\,,\,[0, ty(t)]\big) & x > ty(t) \\ \epsilon_0(A) & x \leq ty(t) \end{cases}.$$

It follows that $K^{(t)}(t, t\,\cdot) \Rightarrow G$ as $t \to \infty$, and additionally that $K^{(t)}(t, \{0\}) \to G(\{0\})$. The relation between $\boldsymbol{X}^{(t)}$ and $\boldsymbol{X}$ is

$$\mathsf{P}_{tu_t}\left[\frac{\boldsymbol{X}_m^{(t)}}{t} \in \cdot \;\middle|\; \tau(t) > m\right] = \mathsf{P}_{tu_t}\left[\frac{\boldsymbol{X}_m}{t} \in \cdot \;\middle|\; \tau(t) > m\right].$$

Theorem 2.3.3. *Let $u_t = u(t) \geq 0$ satisfy $u_t \to u > 0$ as $t \to \infty$. Then on $[0, \infty]^m$, as $t \to \infty$,*

$$\widetilde{\pi}_m^{(t)}(u_t\,,\,\cdot\,) := \mathsf{P}_{tu_t}\left[\left(\frac{X_1^{(t)}}{t}, \ldots, \frac{X_m^{(t)}}{t}\right) \in \cdot\,\right] \Rightarrow \mathsf{P}_u\left[(T_1, \ldots, T_m) \in \cdot\,\right].$$

Thus, the tail chain approxmates $\boldsymbol{X}^{(t)}$ rather than $\boldsymbol{X}$.

Proof. Suppose $m \geq 2$ and $f \in \mathcal{C}[0, \infty]^m$. Then, without loss of generality, $f \in \mathcal{C}_K^+[0, \infty]^m$ as well, since the space is compact. Recall the notation of Theorem

2.3.2. Conditioning on $\tau(t)$, we can write

$$\begin{aligned}
&\widetilde{\pi}_m^{(t)}(u_t\,,\,\cdot)(f)\\
&\quad= \int_{(0,\infty]^m} f(\boldsymbol{x}_m)\, \widetilde{\pi}_m^{(t)}(u_t\,,\, d\boldsymbol{x}_m) + \sum_{k=1}^m \int_{(0,\infty]^{k-1}\times\{0\}^{m-k+1}} f(\boldsymbol{x}_m)\, \widetilde{\pi}_m^{(t)}(u_t\,,\, d\boldsymbol{x}_m)\\
&\quad= \int_{(0,\infty]^m} f(\boldsymbol{x}_m)\, \widetilde{\pi}_m^{(t)}(u_t\,,\, d\boldsymbol{x}_m) + \sum_{k=1}^m \int_{(0,\infty]^{k-1}\times\{0\}} f(\boldsymbol{x}_k,\, 0,\, \ldots,\, 0)\, \widetilde{\pi}_k^{(t)}(u_t\,,\, d\boldsymbol{x}_k)
\end{aligned}$$

by the Markov property. Since

$$\begin{aligned}
\widetilde{\pi}_m^{(t)}(u_t\,,\, \cdot \cap\, (0,\infty]^m) &= \mathsf{P}_{tu_t}\left[t^{-1}\boldsymbol{X}_m^{(t)} \in \cdot\,,\ \tau(t) > m\right]\\
&= \mathsf{P}_{tu_t}\left[t^{-1}\boldsymbol{X}_m \in \cdot \cap (y(t),\infty]^m\right] = \widetilde{\mu}_{m+1}^{(t)}(u_t\,,\, \cdot \times [0,\infty])\,,
\end{aligned}$$

the first term becomes

$$\begin{aligned}
\widetilde{\mu}_{m+1}^{(t)}(u_t\,,\, \cdot \times [0,\infty])(f) &\longrightarrow \mu_{m+1}(u\,,\, \cdot \times [0,\infty])(f)\\
&= \int_{(0,\infty]^m} f(\boldsymbol{x}_m)\, \pi_m(u\,,\, d\boldsymbol{x}_m)\\
&= \int_{(0,\infty]^m} f(\boldsymbol{x}_m)\ \mathsf{P}_u[\boldsymbol{T}_m \in d\boldsymbol{x}_m]
\end{aligned}$$

as $t \to \infty$. Next, for any $A \subset [0,\infty]^k$ measurable, write

$$A_0 = \{\boldsymbol{x}_{k-1} : (\boldsymbol{x}_{k-1}\,,\, 0) \in A\} \subset [0,\infty]^{k-1},$$

and observe that

$$\begin{aligned}
\widetilde{\pi}_k^{(t)}(u_t\,,\, A \cap (0,\infty]^{k-1} \times \{0\}) &= \mathsf{P}_{tu_t}\left[t^{-1}\boldsymbol{X}_{k-1}^{(t)} \in A_0 \cap (0,\infty]^{k-1}\,,\ X_k^{(t)} = 0\right]\\
&= \mathsf{P}_{tu_t}\left[t^{-1}\boldsymbol{X}_{k-1} \in A_0 \cap (y(t),\infty]^{k-1}\,,\ t^{-1}X_k \le y(t)\right]\\
&= \widetilde{\mu}_k^{(t)}(u_t\,,\, A_0 \times [0,\infty]) - \widetilde{\mu}_{k+1}^{(t)}(u_t\,,\, A_0 \times [0,\infty]^2).
\end{aligned}$$

Applying this reasoning to the terms in the summation yields

$$\int_{[0,\infty]^k} f(\boldsymbol{x}_{k-1},0,\ldots,0)\, \widetilde{\mu}_k^{(t)}(u_t\,,\, d\boldsymbol{x}_k) \\ \qquad - \int_{[0,\infty]^{k+1}} f(\boldsymbol{x}_{k-1},0,\ldots,0)\, \widetilde{\mu}_{k+1}^{(t)}(u_t\,,\, d\boldsymbol{x}_{k+1})$$

$$\longrightarrow \int_{[0,\infty]^k} f(\boldsymbol{x}_{k-1},0,\ldots,0)\, \mu_k(u\,,\, d\boldsymbol{x}_k) \\ \qquad - \int_{[0,\infty]^{k+1}} f(\boldsymbol{x}_{k-1},0,\ldots,0)\, \mu_{k+1}(u\,,\, d\boldsymbol{x}_{k+1})$$

$$= \int_{(0,\infty]^{k-1}\times\{0\}} f(\boldsymbol{x}_k,0,\ldots,0)\, \pi_k(u\,,\, d\boldsymbol{x}_k)$$

$$= \int_{(0,\infty]^{k-1}\times\{0\}^{m-k+1}} f(\boldsymbol{x}_m)\ \mathsf{P}_u[\boldsymbol{T}_m \in d\boldsymbol{x}_m].$$

Combining these limits shows that $\mathsf{E}_{tu_t}\, f(t^{-1}\boldsymbol{X}_m^{(t)}) \longrightarrow \mathsf{E}_u\, f(\boldsymbol{T}_m)$ as $t \to \infty$. □

2.4 The Regularity Condition

Previous work on the tail chain derives fdd convergence of $\boldsymbol{X}$ to $\boldsymbol{T}$ under a single assumption analogous to our domain of attraction condition (2.2.5). As we observed in Section 2.3.1, when $G(\{0\}) = 0$, fdd convergence of $\{t^{-1}\boldsymbol{X}\}$ follows directly, but when $G(\{0\}) > 0$, it was common to assume an additional technical condition which made (2.2.5) imply fdd convergence to $\boldsymbol{T}$ as well. This condition, which we refer to as the "regularity condition", controls $\boldsymbol{X}$ upon leaving a non-extreme state. We consider equivalences between different forms appearing in the literature, in terms of both kernels and update functions, and show that, under the regularity condition, the extremal behaviour of $\boldsymbol{X}$ is asymptotically the same as that of its extremal component $\boldsymbol{X}^{(t)}$.

In cases where $G(\{0\}) > 0$, Perfekt [69, 70] requires that

$$\lim_{\delta \downarrow 0} \limsup_{t \to \infty} \sup_{u \in [0,\delta]} K(tu\,,\,(t,\infty]) = 0, \tag{2.4.1}$$

while Segers [84] stipulates that the chosen update function corresponding to K must be of at most linear order in the initial state:

$$\limsup_{t\to\infty} \sup_{0 \le y \le t} t^{-1}\psi(y,v) < \infty, \qquad (v \in B_0,\ \mathsf{P}[V \in B_0] = 1). \tag{2.4.2}$$

Smith [87] used a variant of (2.4.1). We deem a formulation in terms of distributional convergence to be instructive in our context.

Definition. A Markov transition kernel $K \in D(G)$ satisfies the *regularity condition* if

$$K(tu_t\,,\,t\,\cdot) \Rightarrow \epsilon_0(\cdot) \tag{2.4.3}$$

on $[0,\infty]$ as $t \to \infty$ for any non-negative function $u_t = u(t) \to 0$.

Thus, the regularity condition complements the domain of attraction condition expressed in the form (2.2.7) (p. 14), with the effect that the extremal behaviour of K is completely described by the tail kernel K^*.

We now consider the relationships between (2.4.1), (2.4.2) and (2.4.3), and propose an intuitive equivalent for update functions in canonical form.

Proposition 2.4.1. *Suppose $K \in D(G)$, and let $\psi(\cdot, V)$ be an update function corresponding to K such that*

$$t^{-1}\psi(t,v) \longrightarrow \xi(v) \tag{2.4.4}$$

whenever $v \in B$ for which $\mathsf{P}[V \in B] = 1$, and $\xi \circ V \sim G$. Then:

(a) Condition (2.4.1) is necessary and sufficient for K to satisfy the regularity condition (2.4.3).

(b) Condition (2.4.2) *is sufficient for K to satisfy the regularity condition* (2.4.3)*.*

(c) If ψ is in canonical form, i.e.,

$$\psi(y,(Z,W)) = Zy + \phi(y,W),$$

then ψ satisfies (2.4.2) *if and only if $\phi(\cdot, w)$ is bounded on any neighbourhood of* 0 *for each $w \in C$, a set for which* $\mathsf{P}[W \in C] = 1$.

Proof. (a) Assume (2.4.1), and suppose $u_t \to 0$. We show $K(tu_t, t(x,\infty]) \to 0$ for any $x > 0$. Write

$$\omega(t,\delta) = \sup_{u \in [0,\delta]} K(tu\,,\,(t,\infty]).$$

Let $\epsilon > 0$ be given, and choose δ small enough that $\limsup_{t\to\infty} \omega(t,\delta) < \epsilon/2$. Then for t large enough that $u_t < \delta x$, we have

$$K(tu_t\,,\,t(x,\infty]) \le \sup_{u\in[0,\delta x]} K(tu\,,\,t(x,\infty]) = \omega(tx,\delta) < \limsup_{t\to\infty} \omega(t,\delta) + \epsilon/2$$

for t large enough. Our choice of δ implies that $K(tu_t, t(x,\infty]) < \epsilon$.

Conversely, assume that K satisfies (2.4.3) but that (2.4.1) fails. Choose $\epsilon > 0$ and a sequence $\delta_n \downarrow 0$ such that $\limsup_{t\to\infty} \omega(t,\delta_n) \ge \epsilon$ for $n = 1, 2, \ldots$. Then for each n we can find a sequence $t_k^n \to \infty$ as $k \to \infty$ such that $\omega(t_k^n, \delta_n) \ge \epsilon$ for each k. Diagonalize to find $k_1 < k_2 < \cdots$ such that $s_n = t_{k_n}^n \to \infty$ and $\omega(s_n,\delta_n) \ge \epsilon$ for all n. Finally, for $n = 1, 2, \ldots$ choose $u_n \in [0,\delta_n]$ such that

$$K(s_n u_n\,,\,(s_n,\infty]) > \omega(s_n,\delta_n) - \epsilon/2,$$

and put $u(t) = \sum_n u_n \mathbf{1}_{[s_n,s_{n+1})}(t)$. Clearly $u(t) \to 0$, but $K(s_n u(s_n), (s_n,\infty]) \ge \epsilon/2$ for all n, contradicting (2.4.3).

(b) Write $M(v) = \limsup_{t\to\infty} \sup_{0\le y\le t} t^{-1}\psi(y,v)$. Since

$$\sup_{0\le y\le t} t^{-1}\psi(y,v) = \sup_{0\le y\le\delta} \frac{\psi(t\delta^{-1}y,v)}{t\delta^{-1}}\,\delta^{-1}$$

for $\delta > 0$, we have

$$\limsup_{t\to\infty} \sup_{0\le y\le \delta} t^{-1}\psi(ty, v) = \delta M(v).$$

Now, suppose $u_t \to 0$. Given any $\delta > 0$ we have

$$t^{-1}\psi(tu_t, v) \le \sup_{0\le y\le \delta} t^{-1}\psi(ty, v)$$

provided t is large enough, so $\limsup_{t\to\infty} t^{-1}\psi(tu_t, v) \le \delta M(v)$. Consequently,

$$\limsup_{t\to\infty} t^{-1}\psi(tu_t, v) = 0$$

for every v such that $M(v) < \infty$. Under (2.4.2), this means that

$$\mathsf{P}[t^{-1}\psi(tu_t, V) \to 0] = 1,$$

implying (2.4.3).

(c) Suppose first that $\chi_w(a) = \sup_{0\le y\le a} \phi(y, w) < \infty$ for all $a > 0$, whenever $w \in C$. Fixing $w \in C$ and $z \ge 0$, note that

$$\sup_{0\le y\le t} t^{-1}\psi(y, (z, w)) \le z + \sup_{0\le y\le t} t^{-1}\phi(y, w),$$

and observe for any $a > 0$ that

$$\begin{aligned}\sup_{0\le y\le t} t^{-1}\phi(y, w) &\le \Big(\sup_{0\le y\le a} t^{-1}\phi(y, w)\Big) \vee \Big(\sup_{a\le y\le t} y^{-1}\phi(y, w)\Big)\\ &\le t^{-1}\chi_w(a) \vee \Big(\sup_{a\le y} y^{-1}\phi(y, w)\Big).\end{aligned}$$

Choosing a large enough that $\sup_{a\le y} y^{-1}\phi(y, w) \le 1$, say, it follows that

$$\limsup_{t\to\infty} \sup_{0\le y\le t} t^{-1}\psi(y, (z, w)) \le z + 1,$$

so $v = (z, w) \in B_0$. Therefore $\mathsf{P}[(Z, W) \in B_0] \ge \mathsf{P}[Z \ge 0,\ W \in C] = 1$.

Conversely, suppose there is a set D with $\mathsf{P}[W \in D] > 0$ such that $w \in D$ implies $\chi_w(a) = \infty$ for some $0 < a < \infty$. Since $\sup_{0\le y\le t} t^{-1}\psi(y, (z, w)) \ge t^{-1}\chi_w(t)$, we have $[0, \infty) \times D \subset B_0^{\mathrm{c}}$, contradicting (2.4.2). $\square$

The exclusion of necessity from part (b) results from the fact that a kernel K does not uniquely specify an update function ψ. Even when K satisfies the regularity condition (2.4.3), it may be possible to choose a nasty update function ψ which satisfies (2.4.4), but not (2.4.2). However, in such cases there may exist a different update function ψ' corresponding to K which does satisfy (2.4.2).

Here is an example of such a situation. We exhibit an update function ψ for which (i) (2.4.4) holds; (ii) (2.4.2) fails because condition (c) in Proposition 2.4.1 fails; but yet (iii) the corresponding kernel satisfies the regularity condition (2.4.3). Furthermore, we present a different choice of update function corresponding to the same kernel which satisfies (2.4.2). Define $\psi(y, V = (Z, W)) = Zy + \phi(y, W)$, where

$$\phi(y, w) = \sum_{k=1}^{\infty} k \cdot \mathbf{1}_{\{yw=1/k\}}$$

and $W \sim U(0, 1)$. (i) Since $\phi(t, w) = 0$ for $t > 1/w$, it is clear that ψ satisfies (2.4.4) with $\xi = Z$. (ii) Observe that for any $w \in (0, 1)$, $\phi(\cdot, w)$ is unbounded on the interval $[0, 1]$. Therefore, by part (c) of Proposition 2.4.1, (2.4.2) cannot hold for ψ. (iii) However, the corresponding kernel does satisfy the regularity condition (2.4.3). Suppose $u_t \to 0$ and $a > 0$ is arbitrarily large. Write

$$\begin{aligned} \mathsf{P}[t^{-1}\psi(tu_t, (Z, W)) > x] &= \mathsf{P}[Zu_t + t^{-1}\phi(tu_t, W) > x] \\ &\leq \mathsf{P}[t^{-1}\phi(tu_t, W) > x'] + \mathsf{P}[Z > a], \end{aligned}$$

choosing $0 < x' < x - au_t$. Since for any t,

$$\{w : \phi(tu_t, w) > tx'\} \subset \{(tu_t k)^{-1} : k = 1, 2, \ldots\},$$

a set of measure 0 with respect to $\mathsf{P}[W \in \cdot\,]$, (2.4.3) follows by letting $a \to \infty$. On the other hand, the update function $\psi'(y, Z) = Zy$ does satisfy (2.4.2), and for any y,

$$\mathsf{P}\left[\psi'(y, Z) \neq \psi(y, (Z, W))\right] = \mathsf{P}\left[W \in \{(yk)^{-1} : k = 1, 2, \ldots\}\right] = 0,$$

so ψ' does indeed correspond to K.

The regularity condition (2.4.3) restricts attention to Markov chains for which the probability of returning to an extreme state in the next m steps after falling below the extremal boundary is asymptotically negligible. For such chains, as well as those for which $y(t) \equiv 0$ is an extremal boundary for K, the asymptotic behaviour of $\boldsymbol{X}$ is completely accounted for by its extremal component. Hence, the tail chain approximation extends from $\boldsymbol{X}^{(t)}$ to $\boldsymbol{X}$.

Theorem 2.4.1. *Suppose $\boldsymbol{X} \sim K$ with $K \in D(G)$, and let ρ be a metric on $\mathbb{R}^m$. If either*

(i) $y(t) \equiv 0$ is an extremal boundary; or

(ii) K satisfies the regularity condition (2.4.3)*,*

then for any $\epsilon > 0$ we have

$$\mathsf{P}_{tu_t}\left[\rho\Big(\frac{\boldsymbol{X}_m^{(t)}}{t}\,,\,\frac{\boldsymbol{X}_m}{t}\Big) > \epsilon\right] \longrightarrow 0 \qquad (t \to \infty,\ u_t \to u > 0). \tag{2.4.5}$$

Consequently,

$$\mathsf{P}_{tu_t}\left[\left(\frac{X_1}{t}, \ldots, \frac{X_m}{t}\right) \in \cdot\,\right] \Rightarrow \mathsf{P}_u\left[(T_1,\, \ldots,\, T_m) \in \cdot\,\right] \qquad (t \to \infty,\ u_t \to u > 0). \tag{2.4.6}$$

First let us extend the regularity condition to higher-order transition kernels.

Lemma 2.4.1. *If K satisfies* (2.4.3)*, then so do the m-step transition kernels K^m.*

***Proof*.** This is established by induction. Let $u_t \to 0$ and $f \in \mathcal{C}[0,\infty]$. For $m \geq 2$, we have

$$K^m(tu_t\,,\,\cdot)(f) = \int_{[0,\infty]} K^{m-1}(tu_t\,,\,tdv) \int_{[0,\infty]} K(tv\,,\,tdx)\; f(x).$$

Assume that $K^{m-1}(tu_t, t\,\cdot) \Rightarrow \epsilon_0$; (2.4.3) implies that $\int K(tv_t, tdx)f(x) \to f(0)$ whenever $v_t \to 0$. Therefore, by Lemma A.0.2 (a) (p. 132), we conclude that

$$K^m(tu_t\,,\,\cdot)(f) \longrightarrow f(0) = \epsilon_0(f). \qquad \square$$

Proof of Theorem 2.4.1. Suppose $\epsilon > 0$ and $u_t \to u > 0$. Write

$$\mathsf{P}_{tu_t}\left[\rho(t^{-1}\boldsymbol{X}_m^{(t)}\,,\,t^{-1}\boldsymbol{X}_m) > \epsilon\right] = \sum_{k=1}^{m} \mathsf{P}_{tu_t}\left[\rho(t^{-1}\boldsymbol{X}_m^{(t)}\,,\,t^{-1}\boldsymbol{X}_m) > \epsilon\,,\ \tau(t) = k\right].$$

Since $X_j = X_j^{(t)}$ while $j < \tau(t)$, for the k-th summand to converge to 0, it is sufficient that

$$\mathsf{P}_{tu_t}[|X_j^{(t)}/t - X_j/t| > \delta\,,\ \tau(t) = k] = \mathsf{P}_{tu_t}[X_j/t > \delta\,,\ \tau(t) = k] \longrightarrow 0$$

for $j = k, \ldots, m$ and any $\delta > 0$. If $j = k$, we have

$$\mathsf{P}_{tu_t}[X_j/t > \delta\,,\ \tau(t) = k] \leq \mathsf{P}_{tu_t}[X_k/t > \delta,\ X_k/t \leq y(t)] = 0$$

for large t. For $j > k$, recalling the notation of Theorem 2.3.2,

$$\begin{aligned}\mathsf{P}_{tu_t}[X_j/t > \delta\,,\ \tau(t) = k] &= \int_{\mathbb{E}_k'(t)} \mathbf{1}_{[0,y(t)]}(x_k)\ \mathsf{P}_{tu_t}[X_j/t > \delta \,|\, \boldsymbol{X}_k/t = \boldsymbol{x}_k] \\ &\qquad\qquad\qquad\qquad \mathsf{P}_{tu_t}[\boldsymbol{X}_k/t \in d\boldsymbol{x}_k] \\ &= \int_{[0,\infty]^k} \mathsf{P}_{tx_k}[X_{j-k} > t\delta]\,\mathbf{1}_{[0,y(t)]}(x_k)\ \widetilde{\mu}_k^{(t)}(u_t\,,\,d\boldsymbol{x}_k)\end{aligned}$$

using the Markov property. We claim that this integral converges to 0 as $t \to \infty$. If $y(t) \equiv 0$, this follows directly. Otherwise, recall that $\widetilde{\mu}_k^{(t)}(u_t, \cdot) \xrightarrow{v} \mu_k(u, \cdot)$, and consider $h_t(\boldsymbol{x}_k) = \mathsf{P}_{tx_k}[X_{j-k} > t\delta]\,\mathbf{1}_{[0,y(t)]}(x_k)$. Suppose $\boldsymbol{x}^{(t)} \to \boldsymbol{x} \in [0,\infty]^k$. If $x_k > 0$, then $h_t(\boldsymbol{x}^{(t)}) = 0$ for large t because $y(t) \to 0$. Otherwise, if $x_k = 0$, we have $h_t(\boldsymbol{x}^{(t)}) \to 0$ since Lemma 2.4.1 implies that $\mathsf{P}_{tx_k^{(t)}}[X_{j-k} > t\delta] \to 0$ as $t \to \infty$. Lemma A.0.2 (b) establishes (2.4.5); (2.4.6) follows by Slutsky's theorem applied to the result of Theorem 2.3.3. $\square$

Therefore, $\boldsymbol{X}$ converges to $\boldsymbol{T}$ in fdds under (a) $G(\{0\}) = 0$, (b) $G(\{0\}) > 0$ combined with (2.4.3), or (c) $G(\{0\}) > 0$ combined with the extremal boundary $y(t) \equiv 0$. In either case, we will be able to replace the extremal component $\boldsymbol{X}^{(t)}$ with the complete chain $\boldsymbol{X}$ in the results of Sections 3.2.1 and 3.2.2. However, that $y(t) \equiv 0$ is an extremal boundary, and consequently that (2.4.6) holds, does not imply the regularity condition holds, regardless of $G(\{0\})$; in particular, a kernel for which $G(\{0\}) = 0$ need not satisfy (2.4.3). This is illustrated in Example 2.5.3.

2.5 Examples

Our first example illustrates the main results.

Example 2.5.1. Let $V = (Z, \eta)$ be any random vector on $[0, \infty) \times \mathbb{R}$. Consider the update function $\psi(y, V) = (Zy + \eta)_+$ and its canonical form (with $W = (Z, \eta)$)

$$\psi(y, V) = Zy + \phi(y, W) = Zy + \left\{\eta \, \mathbf{1}_{\{\eta > -Zy\}} - Zy \, \mathbf{1}_{\{\eta \le -Zy\}} \right\}.$$

For $y > 0$ and $x \geq 0$, the transition kernel has the form

$$K(y, (x, \infty)) = \mathsf{P}\left[Zy + \eta > x\right].$$

Since $t^{-1}\psi(t, V) = (Z + t^{-1}\eta)_+ \to Z$ a.s., we have $K \in D(G)$ with $G = \mathsf{P}[Z \in \cdot]$. Furthermore, using Proposition 2.3.1, the function $\gamma(t) \equiv \sqrt{t}$ is of larger order than $\phi(t, w)$, so $y(t) = 1/\sqrt{t}$ is an extremal boundary. Since $\phi(\cdot, w)$ is bounded on neighbourhoods of 0, Proposition 2.4.1 (c) implies K satisfies the regularity condition (2.4.3). Consequently, from Theorem 2.4.1, we obtain fdd convergence of $t^{-1}\boldsymbol{X}$ to $\boldsymbol{T}$ as in (2.4.6).

If K does not satisfy the regularity condition (2.4.3), Theorem 2.4.1 may fail to hold, and starting from tu, $t^{-1}\boldsymbol{X}$ may fail to converge to $\boldsymbol{T}$ started from u.

Example 2.5.2. Let $V = (Z, W, W')$ be any non-degenerate random vector on $[0, \infty)^3$, and consider the Markov chain determined by the update function

$$\psi(y, V) = Zy + W\, y^{-1}\, \mathbf{1}_{\{y>0\}} + W'\, \mathbf{1}_{\{y=0\}}\,.$$

For $y > 0$ and $x \geq 0$, the transition kernel is $K(y, (x, \infty)) = \mathsf{P}[Zy + Wy^{-1} > x]$, and since $t^{-1}\psi(t, V) = Z + Wt^{-2} \to Z$ a.s., we have $K \in D(G)$ with $G = \mathsf{P}[Z \in \cdot\,]$. Furthermore, using Proposition 2.3.1, the function $\gamma(t) \equiv 1$ is of larger order than $\phi(t, w)$, so $y(t) = 1/t$ is an extremal boundary.

However, note that $\phi(y, (W, W')) = Wy^{-1}\, \mathbf{1}_{\{y>0\}} + W'\, \mathbf{1}_{\{y=0\}}$ is unbounded near 0, implying that Segers' boundedness condition (2.4.2) does not hold. In fact, our form of the regularity condition (2.4.3) fails for K. Indeed,

$$K(tu_t\,,\, t(x, \infty)) = \mathsf{P}[Ztu_t + W/(tu_t) > tx] = \mathsf{P}[Zu_t + W/(t^2 u_t) > x].$$

Choosing $u_t = t^{-2}$ yields $K(tu_t, t(x, \infty)) \to \mathsf{P}[W > x]$. For appropriate x, this shows (2.4.3) fails.

Not only does (2.4.3) fail but so does Theorem 2.4.1, since the asymptotic behaviour of $\boldsymbol{X}$ is not the same as that of $\boldsymbol{X}^{(t)}$. We show directly that the conditional fdds of $t^{-1}\boldsymbol{X}$ fail to converge to those of $\boldsymbol{T}$. The idea is that if $X_k < y(t) = t^{-1}$, there is a positive probability that $X_{k+1} > t$. We illustrate this for $m = 2$. Take $f \in \mathcal{C}[0, \infty]^2$ and $u > 0$. Observe if $X_0 = tu > 0$, from the definition of ψ, $X_1 = Z_1 tu + W_1/(tu)$ and $X_2 = Z_2 X_1 + (W_2/X_1)\, \mathbf{1}_{\{X_1>0\}} + W_2'\, \mathbf{1}_{\{X_1=0\}}$. Furthermore, on $\{Z_1 > 0\}$, we have $X_1 > 0$ and $X_2 = Z_2 X_1 + W_2/X_1$. On $\{Z_1 = 0, W_1 > 0\}$, $X_1 > 0$ and $X_2 = Z_2 X_1 + W_2/X_1$. On $\{Z_1 = 0, W_1 = 0\}$, we have $X_1 = 0$ and $X_2 = W_2'$. Therefore

$$
\begin{aligned}
&\mathsf{E}_{tu}\, f(X_1/t, X_2/t) \\
&\qquad = \mathsf{E}_{tu}\, f(X_1/t, X_2/t)\, \mathbf{1}_{\{Z_1>0\}} + \mathsf{E}_{tu}\, f(X_1/t, X_2/t)\, \mathbf{1}_{\{Z_1=0, W_1>0\}} \\
&\qquad\qquad + \mathsf{E}_{tu}\, f(X_1/t, X_2/t)\, \mathbf{1}_{\{Z_1=0, W_1=0\}} \\
&\qquad = A + B + C.
\end{aligned}
$$

For A, as $t \to \infty$, we have

$$
\begin{aligned}
A &= \mathsf{E}\, f\big(Z_1 u + W_1/(t^2 u),\ Z_2[Z_1 u + W_1/(t^2 u)] + W_2/[Z_1 t^2 u + W_1 u^{-1}]\big)\, \mathbf{1}_{\{Z_1>0\}} \\
&\longrightarrow \mathsf{E}\, f(Z_1 u, Z_1 Z_2 u)\, \mathbf{1}_{\{Z_1>0\}},
\end{aligned}
$$

while for B we obtain for $t \to \infty$,

$$
\begin{aligned}
B &= \mathsf{E}\, f(W_1/(t^2 u),\ Z_2 W_1/(t^2 u) + W_2 u/W_1)\, \mathbf{1}_{\{Z_1=0, W_1>0\}} \\
&\longrightarrow \mathsf{E}\, f(0, u W_2/W_1)\ \mathbf{1}_{\{Z_1=0, W_1>0\}}\,.
\end{aligned}
$$

Finally for C,

$$
\begin{aligned}
C &= \mathsf{E}\, f(0, W_2'/t)\, \mathbf{1}_{\{Z_1=0, W_1=0\}} = \mathsf{P}[Z_1 = 0, W_1 = 0]\ \mathsf{E}\, f(0, W_2'/t) \\
&\longrightarrow \mathsf{P}[Z_1 = 0, W_1 = 0] f(0, 0).
\end{aligned}
$$

Observe that $\lim_{t\to\infty}[A + B + C] \neq \mathsf{E}_u\, f(T_1, T_2) = \mathsf{E}\, f(uZ_1, uZ_1 Z_2)$.

In the final example, the conditional distributions of $t^{-1}\boldsymbol{X}$ converge to those of the tail chain $\boldsymbol{T}$, even though the regularity condition does not hold. This includes cases for which $G(\{0\}) = 0$ and $G(\{0\}) > 0$ with extremal boundary $y(t) \equiv 0$.

Example 2.5.3. Let $\{(\xi_j, \eta_j) : j \geq 1\}$ be iid copies of the non-degenerate random vector (ξ, η) on $[0, \infty)^2$. Taking $V = (\xi, \eta)$, consider a Markov chain which transitions according to the update function

$$
\psi(y, V) = \xi(y + y^{-1})\, \mathbf{1}_{\{y>0\}} + \eta\, \mathbf{1}_{\{y=0\}} = \xi y + \big\{\xi y^{-1}\, \mathbf{1}_{\{y>0\}} + \eta\, \mathbf{1}_{\{y=0\}}\big\},
$$

where the last expression is the canonical form (with $Z = \xi$ and $W = (\xi, \eta)$). For $y > 0$ and $x \geq 0$, the transition kernel is

$$K(y\,,\,[0,x]) = \mathsf{P}[\xi(y + y^{-1}) \leq x] = \mathsf{P}[\xi \leq x/(y + y^{-1})].$$

For $t > 0$, $t^{-1}\psi(t, V) = \xi(1 + t^{-2}) \to \xi$ a.s., so $K \in D(G)$ with $G = \mathsf{P}[\xi \in \cdot\,]$. Note that $\phi(y, W) = \xi y^{-1}\,\mathbf{1}_{\{y>0\}} + \eta\,\mathbf{1}_{\{y=0\}}$ is unbounded near 0, implying that Segers's boundedness condition (2.4.2) does not hold. Also, our regularity condition (2.4.3) fails for K. To see this, write

$$K(tu_t\,,\,t(x,\infty)) = \mathsf{P}[\xi > x/(u_t + (t^2u_t)^{-1})].$$

Fix x so that $\mathsf{P}[\xi > x] > 0$ and choose $u_t = t^{-2}$. This yields $u_t + (t^2u_t)^{-1} = 1 + t^{-2}$, implying that

$$K(tu_t\,,\,t(x,\infty)) = \mathsf{P}[\xi > x/(1 + t^{-2})] \geq \mathsf{P}[\xi > x] > 0,$$

so (2.4.3) fails for K. However, since $K(t, \{0\}) = \mathsf{P}[\xi = 0] = G(\{0\})$, the choice $y(t) \equiv 0$ satisfies the definition of an extremal boundary (2.3.6), even if $G(\{0\}) > 0$. This leads to fdd convergence of $\mathsf{P}_{tu}[t^{-1}\boldsymbol{X} \in \cdot\,]$ to $\mathsf{P}_u[\boldsymbol{T} \in \cdot\,]$, and thus we learn that the conclusion (2.4.6) of Theorem 2.4.1 may hold without (2.4.3) being true.

2.6 Conclusions and Future Directions

From the update function form, it is apparent that the regularity condition holds in many common applications, meaning that these can be handled using the tail chain approximation as formulated by Perfekt [69, 70] and Segers [84]. However, the introduction of the extremal boundary reinforces the idea that our notion of "extreme" depends on the asymptotic order of states relative to the initial state, and that approximations such as the tail chain are uninformative on asymptotically different orders. This phenomenon is discussed further in the context of the Conditional Extreme Value Model in Chapter 4. Furthermore, our formulation of the regularity condition (2.4.3) (p. 33) in terms of convergence along scaling sequences $u_t \to 0$ makes clear the way in which it complements the basic condition $K \in D(G)$. A natural extension would be to allow convergence along such sequences to a limit distribution H, not necessary ϵ_0, which would describe transitions from non-extreme back to extreme. The theory would then generalize by expanding the definition of the tail kernel K^* to include $K^*(0, A) = H(A)$.

Also, for clarity in defining the extremal boundary, we have restricted ourselves to univariate Markov chains on the non-negative real line. However, the traditional tail chain model has been extended to broader contexts. The ideas developed in this chapter involving the extremal boundary can similarly be generalized, and in so doing will hopefully reveal interesting subtleties in other extreme value models. For example, Segers [84] and Bortot and Coles [19] consider the tail chain on $\mathbb{R}$ as a two-tailed model, which accounts for the phenomenon of "tail-switching", i.e., observing consecutive extremes in opposite tails. A generalization to Markov chains on $\mathbb{R}^d$ would equally cover the case of m-dependent processes; this is explored by Perfekt [70]. Also, the extension by Basrak and Segers [8] to the "tail process" for regularly varying stationary processes suggests that a more general formulation of the extremal boundary may be possible in this setting as well. Moreover, our focus on the asymptotics of transition kernels themselves should prove insightful when extended to the context of general Markov processes.

Two main extensions will be investigated in the following chapters. So far, all of our finite-dimensional results have remained conditional on the initial state.

In Chapter 3, we obtain convergence of the unconditional distributions by adding assumptions on the marginals. This in turn is used to derive the limit of an exceedance point process. Chapter 4 considers a generalization of the condition $K \in D(G)$ where the distribution of the next state is normalized differently from the initial state. However, in this case we are using the transition kernel to describe the dependence structure of a random vector rather than a stochastic process.

CHAPTER 3

CONVERGENCE OF EXTREMAL POINT PROCESSES

3.1 Overview

Point processes have proven a powerful technique for describing the extremal behaviour of stochastic processes. For example, under appropriate conditions on the marginal distributions of a process $\{X_j : j \geq 0\}$, the exceedance point process N_n defined by

$$N_n([0, s] \times [a, \infty]) = \#\{j \leq sn : X_j > ab_n\} \tag{3.1.1}$$

converges weakly to a Poisson limit as $n \to \infty$, where $b_n \to \infty$ is a threshold sequence. From here, one may derive a number of results concerning asymptotic distributions of large order statistics and exceedances of an extreme level. Such results have been pursued in a variety of contexts by Leadbetter et al. [55], Hsing et al. [47, 49, 48], Novak [67], and Balan and Louhichi [5]. More specific results have been obtained for regularly varying processes by Davis and Hsing [29] and Basrak and Segers [8], for regenerative processes by Asmussen [1] and Rootzén [81], and for Markov chains by Perfekt [69] and Yun [89]. Distributions of functionals of such point processes have been considered by Yun [90] and Segers [82, 83].

For stationary processes, the chief consideration is that the dependence structure causes extremes to occur in clusters. The degree of clustering is often described using the extremal index θ, a quantity introduced by Leadbetter et al. [55] which is related to the asymptotic mean cluster size. To obtain a point process convergence result, authors typically employ a mixing condition, such as Leadbetter et al.'s condition $D(u_n)$ [55, p. 53], to split the process into approximately independent and identically distributed blocks. With an appropriate choice of block size, extremes within one such block belong asymptotically to the same cluster. Under an assumption controlling the extremal behaviour within each block, such as on the distribution of the number of exceedances, N_n generally converges to a limiting compound Poisson process, where the compounding at each timepoint approximates the clustering within separate blocks.

For Markov chains, the behaviour upon reaching an extreme state can be modelled by the tail chain, as we saw in Chapter 2. Point process results for stationary Markov chains employ the tail chain to specify the compounding in the limit process. Under Markov dependence, the within-block behaviour is determined merely by conditions on the marginal distribution and the transition kernel. Basrak and Segers [8] have since extended the tail chain model to general jointly regularly varying stationary processes.

Taking a somewhat different approach, Rootzén [81] directs attention to regenerative processes, which split naturally into "cycles". In this case, the within-block condition is replaced by an assumption on the extremal behaviour over a cycle. The main difference is that the cycles are of random but finite length, whereas the block size increases deterministically with n. In particular, Rootzén shows that the limit of the process counting the number of exceedances depends on the asymptotics of the distribution of the cycle maximum, as well as the marginal distribution.

We combine these two approaches to derive the weak limit of N_n when $\{X_n\}$ is a positive recurrent Markov chain. Such chains display a regenerative structure, with regenerations occuring upon visits to a recurrent set. In the limit, N_n is approximated by a process consisting of clusters of points stacked above common timepoints, each corresponding to a separate regenerative cycle. The heights of the points in each cluster are determined by an independent run of the tail chain. We follow the theoretical development of the tail chain process outlined in Chapter 2, applied to a Markov chain with heavy-tailed marginals. The results in Chapter 4, albeit couched in a different context, suggest a way this theory can be extended to accomodate more general choices of marginal distribution.

3.2 Regular Variation Properties of the Extremal Component

Recall $\boldsymbol{X} = (X_0, X_1, \dots)$ is a Markov chain on $[0, \infty)$ with transition kernel $K \in D(G)$. We first discuss a sense in which the extremal component $\boldsymbol{X}^{(t)}$ can be thought of as jointly regularly varying, with limit measure determined by the tail chain. This will play a fundamental role in deriving the limit of N_n, given by (3.1.1).

So far our convergence results required that the initial state become extreme. To obtain a result for the unconditional distribution of $(X_0, \dots, X_m)$, we require additional assumptions about how likely the individual observations X_j are to be large.

3.2.1 Effect of a Regularly Varying Initial Distribution

When the distribution of X_0 has a regularly varying tail, the results of the previous sections extend to regular variation on the cone $\mathbb{E}_m = (0, \infty] \times [0, \infty]^m$ using Lemma A.0.4. This can be interpreted as a result conditional on the first observation exceeding a threshold. The cone $\mathbb{E}_m$ is smaller than the cone $[0,\infty]^{m+1}\backslash\{\mathbf{0}\}$ traditionally employed in extreme value theory, reflecting the fact that this extension remains uninformative when the initial state is not extreme. Regular variation on the cone $\mathbb{E}_m$ is analogous to the Conditional Extreme Value Model for a random vector discussed in Chapter 4. The basics of regular variation are discussed in Section 1.2.2 (p. 6).

Proposition 3.2.1. *Assume $\boldsymbol{X} \sim K \in D(G)$ is a Markov chain on $[0,\infty)$, and $X_0 \sim F$ with $1-F \in \mathrm{RV}_{-\alpha}$, i.e., $tF(b(t)\,\cdot) \xrightarrow{v} \nu_\alpha$ in $\mathbb{M}_+(0,\infty]$ as $t\to\infty$, where $b(t)\to\infty$ (see Section 1.2.2). On $\mathbb{E}_m = (0,\infty]\times[0,\infty]^m$, define the measure*

$$\mu(dx_0, d\boldsymbol{x}_m) = \nu_\alpha(dx_0)\,\mathsf{P}_{x_0}\big[(T_1,\,\ldots,\,T_m)\in d\boldsymbol{x}_m\big]. \tag{3.2.1}$$

Then, for $m=1,2,\ldots$, the following convergences take place as $t\to\infty$:

(a) In $\mathbb{M}_+((0,\infty]^m\times[0,\infty])$,

$$t\,\mathsf{P}\left[\left(\frac{X_0}{b(t)},\ldots,\frac{X_m}{b(t)}\right)\in\cdot\cap(0,\infty]^m\times[0,\infty]\right]\xrightarrow{v}\mu\big(\cdot\cap(0,\infty]^m\times[0,\infty]\big).$$

(b) In $\mathbb{M}_+(\mathbb{E}_m)$,

$$t\,\mathsf{P}\left[\left(\frac{X_0^{(b(t))}}{b(t)},\ldots,\frac{X_m^{(b(t))}}{b(t)}\right)\in\cdot\right]\xrightarrow{v}\mu(\cdot). \tag{3.2.2}$$

(c) If either (i) $G(\{0\})=0$; (ii) $y(t)\equiv 0$ is an extremal boundary; or (iii) K satisfies the regularity condition (2.4.3), then in $\mathbb{M}_+(\mathbb{E}_m)$,

$$t\,\mathsf{P}\left[\left(\frac{X_0}{b(t)},\ldots,\frac{X_m}{b(t)}\right)\in\cdot\right]\xrightarrow{v}\mu(\cdot).$$

(d) In $\mathbb{M}_+(0,\infty]$,

$$t\,\mathsf{P}\left[\frac{X_0}{b(t)}\in dx_0\,,\ \tau(b(t))\geq m\right]\xrightarrow{v}\big[1-G(\{0\})\big]^{m-1}\nu_\alpha(dx_0).$$

Note that limit measure μ is required to be finite on relatively compact subsets of $\mathbb{E}_m$, which are contained in rectangles of the form $[x_0, \infty] \times [0, \infty]^m$.

Remark. These convergence statements may be reformulated equivalently as, say,

$$\mathsf{P}\left[\left(\frac{X_0^{(b(t))}}{b(t)}, \dots, \frac{X_m^{(b(t))}}{b(t)}\right) \in \cdot \;\middle|\; X_0 > \delta b(t)\right] \Rightarrow \mathsf{P}\left[(T_0, \dots, T_m) \in \cdot\right],$$

where $T_0 \sim \text{Pareto}(\alpha)$ supported on (δ, ∞). This is the form considered by Segers [84]. In this sense, these results are conditional on the first observation being extreme. Note that, up to this point, we have not imposed any restrictions on the distribution G.

Proof. Apply Lemma A.0.4 (p. 133) to the results of Theorems 2.3.1 (p. 20), 2.3.3 (p. 30) and 2.4.1 (p. 37), and to (2.3.15) (p. 29). □

For convenience in what follows, we provide formulas for a two-dimensional measure similar to μ.

Lemma 3.2.1. *For $\alpha > 0$ and a random variable $Y \geq 0$ with $\mathsf{E}\, Y^\alpha \leq \infty$, consider the measure*

$$\nu(du, dv) = \nu_\alpha(du)\, \mathsf{P}[uY \in dv] \qquad \textit{on } \mathbb{E}_1 = (0, \infty] \times [0, \infty].$$

We have

$$\nu((x, \infty] \times [0, y]) = x^{-\alpha}\, \mathsf{P}[Y \leq yx^{-1}] - y^{-\alpha}\, \mathsf{E}[Y^\alpha\, \mathbf{1}_{\{Y \leq yx^{-1}\}}] < \infty;$$

$$\nu((0, x] \times (y, \infty]) = y^{-\alpha}\, \mathsf{E}[Y^\alpha\, \mathbf{1}_{\{Y > yx^{-1}\}}] - x^{-\alpha}\, \mathsf{P}[Y > yx^{-1}] \leq \infty; \tag{3.2.3}$$

$$\nu((0, \infty] \times (y, \infty]) = y^{-\alpha}\, \mathsf{E}\, Y^\alpha \leq \infty. \tag{3.2.4}$$

The last two are finite if and only if $\mathsf{E}\, Y^\alpha < \infty$.

***Proof*.** Observe that

$$\nu((x,\infty]\times[0,y]) = \int_{(x,\infty]} \nu_\alpha(du)\,\mathsf{P}[Y \le yu^{-1}] = \int_{[0,x^{-\alpha}]} du\,\mathsf{P}[Y \le yu^{1/\alpha}]$$
$$= \int_{[0,x^{-\alpha}]} du\,\mathsf{P}[y^{-\alpha}Y^\alpha \le u]$$

by change of variables. Applying Fubini's theorem, this becomes

$$\int_{[0,x^{-\alpha}]} (x^{-\alpha}-u)\,\mathsf{P}[y^{-\alpha}Y^\alpha \in du] = x^{-\alpha}\,\mathsf{P}[Y \le yx^{-1}] - y^{-\alpha}\,\mathsf{E}[Y^\alpha\,\mathbf{1}_{\{Y\le yx^{-1}\}}].$$

Now, $\nu((x,\infty]\times(y,\infty]) = x^{-\alpha} - \nu((x,\infty]\times[0,y])$. Letting $x\to\infty$ yields (3.2.4) by monotone convergence, whether $\mathsf{E}\,Y^\alpha$ is finite or not. Finally, $\nu((0,x]\times(y,\infty]) = \nu((0,\infty]\times(y,\infty]) - \nu((x,\infty]\times(y,\infty])$. □

Returning to Proposition 3.2.1, note that in the case $m=1$, the cone $\mathbb{E}_1$ is a rotated version of the space $[0,\infty]\times(0,\infty]$ used in the Conditional Extreme Value odel (see Chapter 4). The limit can be expressed as

$$\mu((x_0,\infty]\times[0,x_1]) = x_0^{-\alpha}\,\mathsf{P}[\xi \le x_1/x_0] - x_1^{-\alpha}\,\mathsf{E}\,\xi^\alpha\,\mathbf{1}_{\{\xi\le x_1/x_0\}},$$

where $\xi \sim G$. In particular,

$$\mu((x_0,\infty]\times(0,\infty]) = x_0^{-\alpha}\,\mathsf{P}[\xi>0] \qquad \text{and} \qquad \mu((0,\infty]\times(x_1,\infty]) = x_1^{-\alpha}\,\mathsf{E}\,\xi^\alpha.$$

Thus, the quantities $G(\{0\})$ and $\mathsf{E}\,\xi^\alpha$ parametrize the mass assigned by μ to certain slices of the space $\mathbb{E}_m$.

3.2.2 Joint Tail Convergence

What additional assumptions are necessary for convergences (b) and (c) of the previous result to take place on the larger cone $\mathbb{E}_m^* = [0,\infty]^{m+1}\backslash\{\mathbf{0}\}$? This was considered by Segers [84] and Basrak and Segers [8] for stationary Markov chains. In (b), the dependence on the extremal threshold and hence on t means we are in the context of a triangular array and are not, strictly speaking, dealing with joint regular variation. However, the result will still be useful in deriving a point process convergence via the Poisson transform [77, p. 183].

As a first step, we characterize convergence on the larger cone by decomposing it into smaller, more familiar cones. This is similar to [84, Theorem 6.1] and one of the implications of [8, Theorem 2.1]. As a convention in what follows, set $[0,\infty]^0 \times A = A$. Also, recall the notation $\mathbb{E}_m = (0,\infty] \times [0,\infty]^m$.

Proposition 3.2.2. *Suppose for $t > 0$ that $\boldsymbol{Y}_t = (Y_{t,0}, Y_{t,1}, \dots, Y_{t,m})$ is a random vector on $[0,\infty]^{m+1}$. There exists a non-null Radon measure μ^* on $\mathbb{E}_m^* = [0,\infty]^{m+1} \backslash \{\mathbf{0}\}$ such that*

$$t\,\mathsf{P}\left[(Y_{t,0}, Y_{t,1}, \dots, Y_{t,m}) \in \cdot\,\right] \xrightarrow{v} \mu^*(\cdot) \qquad \text{in } \mathbb{M}_+(\mathbb{E}_m^*) \qquad (t \to \infty) \tag{3.2.5}$$

iff for $j = 0, \dots, m$ there exist Radon measures μ_j on $\mathbb{E}_j = (0,\infty] \times [0,\infty]^j$, not all null, such that

$$t\,\mathsf{P}\left[(Y_{t,j}, \dots, Y_{t,m}) \in \cdot\,\right] \xrightarrow{v} \mu_{m-j}(\cdot) \qquad \text{in } \mathbb{M}_+(\mathbb{E}_{m-j}). \tag{3.2.6}$$

The relation between the limit measures is the following:

$$\mu_{m-j}(\cdot) = \mu^*([0,\infty]^j \times \cdot\,) \quad \text{on } \mathbb{E}_{m-j}$$

for $j = 0, \dots, m$, and

$$\mu^*\big([\mathbf{0}, \boldsymbol{x}]^{\mathrm{c}}\big) = \sum_{j=0}^{m} \mu_{m-j}\big((x_j, \infty] \times [0, x_{j+1}] \times \cdots \times [0, x_m]\big) \quad \text{for } \boldsymbol{x} \in \mathbb{E}_m^*.$$

Furthermore, given $j \in \{0, \dots, m-1\}$, if $A \subset [0,\infty]^{m-j} \backslash \{0\}^{m-j}$ is relatively compact, then $\mu_{m-j}((0,\infty] \times A) < \infty$.

Proof. Assume first that (3.2.5) holds. Fixing $j \in \{0, \dots, m\}$, define $\mu_{m-j}(\cdot) := \mu^*([0,\infty]^j \times \cdot)$ (i.e., $\mu_m = \mu^*$). Let $A \subset \mathbb{E}_{m-j}$ be relatively compact with $\mu_{m-j}(\partial A) = 0$. Then $A^* = [0,\infty]^j \times A$ is relatively compact in $\mathbb{E}_m^*$, and $\partial_{\mathbb{E}_m^*} A^* = [0,\infty]^j \times \partial_{\mathbb{E}_{m-j}} A$, so $\mu^*(\partial_{\mathbb{E}_m^*} A^*) = \mu_{m-j}(\partial A) = 0$. Therefore,

$$t\,\mathsf{P}\left[(Y_{t,j}, \dots, Y_{t,m}) \in A\right] = t\,\mathsf{P}\left[(Y_{t,0}, \dots, Y_{t,m}) \in A^*\right] \longrightarrow \mu^*(A^*) = \mu_{m-j}(A),$$

establishing (3.2.6). Conversely, suppose we have (3.2.6) for $j = 0, \dots, m$. For $\boldsymbol{x} \in (0,\infty]^{m+1}$, define

$$h(\boldsymbol{x}) = \sum_{j=0}^{m} \mu_{m-j}\big((x_j, \infty] \times [0, x_{j+1}] \times \cdots \times [0, x_m]\big).$$

Decompose $[\mathbf{0}, \boldsymbol{x}]^{\mathrm{c}}$ as a disjoint union

$$[\mathbf{0}, \boldsymbol{x}]^{\mathrm{c}} = \bigcup_{j=0}^{m} [0,\infty]^j \times (x_j, \infty] \times [0, x_{j+1}] \times \cdots \times [0, x_m]\,, \tag{3.2.7}$$

and observe that at points of continuity of the limit,

$$\begin{aligned} &t\,\mathsf{P}\left[\boldsymbol{Y}_t \in [\mathbf{0}, \boldsymbol{x}]^{\mathrm{c}}\right] \qquad (3.2.8)\\ &\quad= \sum_{j=0}^{m} t\,\mathsf{P}\left[(Y_{t,j}, \dots, Y_{t,m}) \in (x_j, \infty] \times [0, x_{j+1}] \times \cdots \times [0, x_m]\right] \longrightarrow h(\boldsymbol{x}). \end{aligned}$$

It follows that (3.2.5) holds with the limit measure μ^* defined by $\mu^*([\mathbf{0}, \boldsymbol{x}]^{\mathrm{c}}) = h(\boldsymbol{x})$. Indeed, given $f \in \mathcal{C}_K^+(\mathbb{E}_m^*)$ we can find $\delta > 0$ such that $\boldsymbol{x}_\delta = (\delta, \dots, \delta)$ is a continuity point of h and f is supported on $[\mathbf{0}, \boldsymbol{x}_\delta]^{\mathrm{c}}$. Therefore,

$$t\,\mathsf{E}\,f(\boldsymbol{Y}_t) \le \sup_{\boldsymbol{x} \in \mathbb{E}_m^*} f(\boldsymbol{x}) \cdot \sup_{t>0} t\,\mathsf{P}\left[\boldsymbol{Y}_t \in [\mathbf{0}, \boldsymbol{x}_\delta]^{\mathrm{c}}\right] < \infty,$$

implying that the set $\{t\,\mathsf{P}[\boldsymbol{Y}_t \in \cdot\,]\,;\, t > 0\}$ is relatively compact in $\mathbb{M}_+(\mathbb{E}_m^*)$ (see Resnick [77, p. 51]). Furthermore, if $t_k\,\mathsf{P}[\boldsymbol{Y}_{t_k} \in \cdot\,] \to \mu$ and $s_k\,\mathsf{P}[\boldsymbol{Y}_{s_k} \in \cdot\,] \to \mu'$ as $k \to \infty$, then $\mu = \mu' = \mu^*$ on sets $[\mathbf{0}, \boldsymbol{x}]^{\mathrm{c}}$ which are continuity sets of μ^* by (3.2.8). This extends to measurable rectangles in $\mathbb{E}_m^*$ bounded away from $\mathbf{0}$ whose vertices are continuity points of h, leading us to the conclusion that $\mu =$

$\mu' = \mu^*$ on $\mathbb{E}_m^*$. Moreover, since we can decompose $[\mathbf{0}, \boldsymbol{x}]^{\mathrm{c}}$ for any $\boldsymbol{x} \in \mathbb{E}_m^*$ as in (3.2.7), it is clear that μ^* is non-null if and only if not all of the μ_j are null. Finally, for $1 \le j \le m-1$, if $A \subset [0,\infty]^{m-j} \backslash \{0\}^{m-j}$ is relatively compact, then it is contained in $[(0,\ldots,0),(x_{j+1},\ldots,x_m)]^{\mathrm{c}}$ for some $(x_{j+1},\ldots,x_m) \in (0,\infty]^{m-j}$. Applying (3.2.7) once again, we find that

$$\begin{aligned}\mu_{m-j}((0,\infty] \times A) &= \mu^*([0,\infty]^j \times (0,\infty] \times A) \\ &\le \sum_{k=j+1}^{m} \mu^*\big([0,\infty]^{j+1} \times [0,\infty]^{k-j-1} \times (x_k,\infty] \times [0,x_{k+1}] \times \cdots \times [0,x_m]\big) \\ &= \sum_{k=j+1}^{m} \mu_{m-k}\big((x_k,\infty] \times [0,x_{k+1}] \times \cdots \times [0,x_m]\big) < \infty. \qquad \square\end{aligned}$$

Consequently, due to the Markovian structure, the extension of the convergences in Proposition 3.2.1 to the larger cone $\mathbb{E}_m^*$ follows from regular variation of the marginal tails, as well as a moment condition on G.

Theorem 3.2.1. *Suppose $\boldsymbol{X} \sim K \in D(G)$, and $X_0 \sim F$ satisfying $tF(b(t)\cdot) \xrightarrow{v} \nu_\alpha$ in $\mathbb{M}_+(0,\infty]$, where $b(t) \to \infty$. Let $\xi \sim G$. Then*

$$t\,\mathsf{P}\left[\left(\frac{X_0^{(b(t))}}{b(t)},\ldots,\frac{X_m^{(b(t))}}{b(t)}\right) \in \cdot\,\right] \xrightarrow{v} \mu^*(\cdot) \qquad \text{in } \mathbb{M}_+(\mathbb{E}_m^*) \qquad (t \to \infty), \quad (3.2.9)$$

where μ^* is the Radon measure on $\mathbb{E}_m^*$ given by

$$\mu^*\big|_{\mathbb{E}_m}(dx_0, d\boldsymbol{x}_m) = \nu_\alpha(dx_0)\,\mathsf{P}_{x_0}\left[(T_1,\ldots,T_m) \in d\boldsymbol{x}_m\right] \quad \text{and} \quad \mu^*(\mathbb{E}_m^* \backslash \mathbb{E}_m) = 0,$$

if and only if $\mathsf{E}\,\xi^\alpha < \infty$ and

$$t\,\mathsf{P}\left[X_j^{(b(t))}/b(t) \in \cdot\,\right] \xrightarrow{v} (\mathsf{E}\,\xi^\alpha)^j\,\nu_\alpha(\cdot) \quad \text{in } \mathbb{M}_+(0,\infty], \quad j = 1,\ldots,m. \quad (3.2.10)$$

The Radon property of μ^* on $\mathbb{E}_m^*$ requires that sets of the form $[0,x_0] \times [\mathbf{0},\boldsymbol{x}_m]^{\mathrm{c}}$ now have finite measure. Although the convergence (3.2.9) is not, strictly speaking, joint regular variation of a random vector, since the distributions $\mathsf{P}[\boldsymbol{X}_m^{(b(t))} \in \cdot\,]$ are indexed by t, it is convenient to draw the analogy and to refer to the property (3.2.9) holding for $m \geq 0$ as "joint regular variation of the extremal component".

Proof. Assume first that (3.2.9) holds, and let $x > 0$. For $j \geq 1$, (3.2.9) implies

$$t \, \mathsf{P}\left[X_j^{(b(t))} > b(t)x\right] \longrightarrow \mu^*([0,\infty]^j \times (x,\infty] \times [0,\infty]^{m-j})$$
$$= \int_{(0,\infty]} \nu_\alpha(du) \; \mathsf{P}\left[\xi_1 \cdots \xi_j > xu^{-1}\right] = x^{-\alpha}\, \mathsf{E}(\xi_1 \cdots \xi_j)^\alpha = x^{-\alpha}(\mathsf{E}\,\xi^\alpha)^j < \infty$$

by (3.2.4), since the set $[0,\infty]^j \times (x,\infty] \times [0,\infty]^{m-j}$ is relatively compact in $\mathbb{E}_m^*$. In particular, we have $\mathsf{E}\,\xi^\alpha < \infty$. Conversely, suppose that (3.2.10) holds. Lemma A.0.4 implies that in $\mathbb{M}_+(\mathbb{E}_{m-j})$,

$$t \, \mathsf{P}\left[b(t)^{-1}(X_j^{(b(t))}, \ldots, X_m^{(b(t))}) \in (dx_0, d\boldsymbol{x})\,\right]$$
$$\xrightarrow{v} (\mathsf{E}\,\xi^\alpha)^j \nu_\alpha(dx_0) \; \mathsf{P}_{x_0}\left[(T_1, \ldots, T_{m-j}) \in d\boldsymbol{x}\right] =: \mu_{m-j}(dx_0, d\boldsymbol{x})$$

by the Markov property, and Proposition 3.2.2 yields (3.2.9), with $\mu^*|_{\mathbb{E}_m}(\cdot) = \mu_m(\cdot)$. It remains to verify that $\mu^*(\mathbb{E}_m^* \backslash \mathbb{E}_m) = 0$. Writing $A_j(\boldsymbol{x}_m) = [0,\infty]^{j-1} \times (x_j,\infty] \times [0,\infty]^{m-j}$, note that

$$\mu^*(\{0\} \times [\boldsymbol{0}, \boldsymbol{x}_m]^c) \leq \sum_{j=1}^m \mu^*(\{0\} \times A_j(\boldsymbol{x}_m)),$$

and

$$\begin{aligned}\mu^*(\{0\} \times A_j(\boldsymbol{x}_m)) &= \mu_{m-j}((x_j,\infty] \times [0,\infty]^{m-j}) - \mu_m((0,\infty] \times A_j(\boldsymbol{x}_m)) \\ &= (\mathsf{E}\,\xi^\alpha)^j x_j^{-\alpha} - \lim_{x_0 \downarrow 0} \int_{[x_0,\infty]} \nu_\alpha(du) \; \mathsf{P}\left[\xi_1 \cdots \xi_j > x_j u^{-1}\right] \\ &= (\mathsf{E}\,\xi^\alpha)^j x_j^{-\alpha} - (\mathsf{E}\,\xi^\alpha)^j x_j^{-\alpha} = 0.\end{aligned}$$

Since $\mathbb{E}_m^* \backslash \mathbb{E}_m = \{0\} \times \mathbb{E}_{m-1}^*$, it follows that $\mu^*(\mathbb{E}_m^* \backslash \mathbb{E}_m) = 0$. □

Regardless of whether the convergence (3.2.10) takes place, the limit always constitutes a lower bound on the tail weight of $X_j^{(b(t))}$, since

$$\liminf_{t\to\infty} t \, \mathsf{P}\left[\frac{X_j^{(b(t))}}{b(t)} > x\right] \geq \mu\big((0,\infty] \times [0,\infty]^{j-1} \times (x,\infty] \times [0,\infty]^{m-j}\big) = (\mathsf{E}\,\xi^{\,\alpha})^j x^{-\alpha}$$

by (3.2.2), applying (3.2.4). In fact, (3.2.10) amounts to moment conditions on the $X_j^{(b(t))}$, since under (3.2.2), (3.2.10) holds if there exists $\varepsilon > 0$ such that

$$\lim_{\delta\downarrow 0}\limsup_{t\to\infty} t\,\mathsf{E}\left[\left(\frac{X_j^{(b(t))}}{b(t)}\right)^{\varepsilon}\mathbf{1}_{\{X_0\le\delta b(t)\}}\right] = 0 \qquad j = 1,\ldots,m \tag{3.2.11}$$

by Markov's inequality. See Maulik et al. [61] for details involving an analogous condition.

The relationship between $X_j^{(b(t))}$ and X_j is given by

$$\mathsf{P}\left[\frac{X_j^{(b(t))}}{b(t)} > x\right] = \mathsf{P}\left[\frac{X_0}{b(t)} > y(b(t)),\ldots,\frac{X_{j-1}}{b(t)} > y(b(t)),\frac{X_j}{b(t)} > x\right] \le \mathsf{P}\left[\frac{X_j}{b(t)} > x\right],$$

implying that the tail weight of X_j bounds that of $X_j^{(b(t))}$. Therefore, knowledge concerning the tail behaviour of X_j imposes a restriction on the distributions G to which K can be attracted via the α-th moment. For example, if it is known that $t\,\mathsf{P}[X_1/b(t)\in\cdot] \xrightarrow{v} \nu_\alpha$, such as when $\boldsymbol{X}$ is stationary, then we must have $\mathsf{E}\,\xi^\alpha \le 1$.

At the end of Section 2.4, cases were outlined in which we could replace $X_j^{(b(t))}$ by X_j. Theorem 3.2.1 is most striking for these since it shows that for a Markov chain whose kernel is in a domain of attraction, to obtain joint regular variation of the fdds it is enough to know that the marginal tails are regularly varying. In particular, if $\boldsymbol{X}$ has a regularly varying stationary distribution then the fdds are jointly regularly varying. This result was presented by Segers [84], and Basrak and Segers [8] showed that for a general stationary process, joint regular variation of fdds is equivalent to the existence of a "tail process" which reduces to the tail chain in the case of Markov chains. However, what Proposition 3.2.1 emphasizes is that it is the marginal tail behaviour alone, rather than stationarity, which provides the link with joint regular variation.

3.2.3 Maximum of the Extremal Component

We now give some conditions on the extremal component which will lead to a meaningful point process limit. These allow us to deal with the positive portion of the extremal component, the random vector of random length

$$\{X_j^{(t)} : j = 0, \ldots, \tau(t) - 1\} = \{X_j : j = 0, \ldots, \tau(t) - 1\},$$

rather than the finite-dimensional projections $\{X_j^{(t)} : j = 0, \ldots, m\}$. This will lead to further restrictions on the behaviour of the tail chain $\boldsymbol{T}$.

We study a positive recurrent chain $\boldsymbol{X}$ by splitting it into regenerative cycles and analyzing its extremal properties via the extremal components of the cycles. Asmussen [1] and Rootzén [81] point out that, for regenerative processes, there is a close connection between point process convergence and the asymptotic distribution of the cycle maximum. Following this approach, we investigate conditions under which the distribution of the maximum over the extremal component is regularly varying in a certain sense. The limit measure in this result is useful in computing the extremal index for the process $\boldsymbol{X}$ (see Rootzén [81]).

Condition 3.2.1.

$$\lim_{m\to\infty} \limsup_{t\to\infty} \mathsf{P}\left[\sup_{j\geq m} \frac{X_j^{(b(t))}}{b(t)} > a \,\middle|\, X_0 > \delta b(t)\right] = 0 \qquad \text{for all } a, \delta > 0.$$

This condition controls the persistence of non-zero values of the extremal component. It is the counterpart of [8, Condition 4.1] and [69, (3.1)], which are formulated in terms of block sizes; indeed,

$$\sup_{j\geq m} X_j^{(t)} = \left(\sup_{m\leq j<\tau(t)} X_j\right) \mathbf{1}_{\{\tau(t)>m\}}\,.$$

It can be thought of as a tightness condition to complement the finite-dimensional convergences (3.2.2). Simpler forms exist depending on whether $G(\{0\}) > 0$ or $G(\{0\}) = 0$; see Section 3.3.

Another pertinent interpretation relates to the stability of the Markov chain $\boldsymbol{X}$. Condition 3.2.1 essentially requires that the chain drifts back to the non-extreme states after visiting an extreme state. In other words, the non-extreme states are recurrent. This property is reflected in the tail chain, which is transient under Condition 3.2.1.

Recalling the tail chain expressed as in (2.2.4), we introduce the notation $\xi(n) = \prod_{j=1}^{n} \xi_j$, so that $T_n = T_0 \cdot \xi(n)$.

Proposition 3.2.3. *Let* $\boldsymbol{X} \sim K \in D(G)$ *be a Markov chain on* $[0, \infty)$ *with initial distribution* F *satisfying* $1 - F \in \mathrm{RV}_{-\alpha}$*, so that* (3.2.2) *(p. 48) holds. Suppose that* $\boldsymbol{X}$ *satisfies Condition 3.2.1. Then*

$$\lim_{m\to\infty} \mathsf{P}\left[\sup_{j\geq m} \xi(j) > a\right] = 0, \qquad a > 0. \tag{3.2.12}$$

Remark. By a standard argument, (3.2.12) implies that the tail chain is a transient random walk, i.e.,

$$\mathsf{P}[T_m \to 0] = 1. \tag{3.2.13}$$

It is important to note that, even though the tail chain $\boldsymbol{T}$ lives on the same state space as $\boldsymbol{X}$, namely $[0, \infty)$, the interpretation of the states differs between the two. Indeed, for $\boldsymbol{T}$, $\{0\}$ is a special boundary state which represents the collection of non-extreme states of $\boldsymbol{X}$ under the tail chain approximation. If we consider instead the (additive) random walk $\{\log T_n\}_{n\geq 0}$, (3.2.13) becomes $\mathsf{P}[\log T_m \to -\infty] = 1$.

Proof. Observe that

$$
\begin{aligned}
t\,\mathsf{P}\Big[X_0 > b(t),\ \sup_{m\le j\le r} X_j^{(b(t))} \le b(t)\Big]& \\
&= t\,\mathsf{P}\big[X_0 > b(t),\ X_m^{(b(t))} \le b(t),\ \dots,\ X_r^{(b(t))} \le b(t)\big] \\
&\longrightarrow \int_{(1,\infty]} \nu_\alpha(dx)\,\mathsf{P}_x\big[T_m \le 1,\ \dots,\ T_r \le 1\big] \\
&= \int_{(1,\infty]} \nu_\alpha(dx)\,\mathsf{P}\Big[\sup_{m\le j\le r} \xi(j) \le x^{-1}\Big]
\end{aligned}
$$

by (3.2.2). Therefore,

$$
\begin{aligned}
\limsup_{t\to\infty} t\,\mathsf{P}\Big[X_0 > b(t),\ \sup_{j\ge m} X_j^{(b(t))} > b(t)\Big] &\ge \lim_{r\to\infty} \int_{(1,\infty]} \nu_\alpha(dx)\,\mathsf{P}\Big[\sup_{m\le j\le r} \xi(j) > x^{-1}\Big] \\
= \int_{(1,\infty]} \nu_\alpha(dx)\,\mathsf{P}\Big[\sup_{j\ge m} \xi(j) > x^{-1}\Big] &=: \int_{(1,\infty]} \nu_\alpha(dx) f_m(x).
\end{aligned}
$$

Now, Condition 3.2.1 implies that $\int_{(1,\infty]} \nu_\alpha(dx) f_m(x) \to 0$ as $m \to \infty$. We claim that $f_m(x) \to 0$ for any $x > 0$. Suppose instead that $\inf_m f_m(x_0) \ge c > 0$ for some x_0. Since the f_m are all increasing in x, we have $\inf_m f_m(x) \ge c$ for $x \ge x_0$. But this implies that

$$
\liminf_{m\to\infty} \int_{(1,\infty]} \nu_\alpha(dx) f_m(x) \ge \liminf_{m\to\infty} \int_{(1\vee x_0,\infty]} \nu_\alpha(dx) f_m(x) \ge c\nu_\alpha(1\vee x_0, \infty] > 0
$$

by Fatou's Lemma, contradicting Condition 3.2.1. Therefore, $f_m(x) \to 0$ for all $x > 0$ as $m \to \infty$, establishing (3.2.12). □

Condition 3.2.1 is conditional on the first observation exceeding $\delta b(t)$, similar to (3.2.2). In situations where the stronger convergence of unconditional distributions (3.2.9) takes place, we will require an additional assumption.

Condition 3.2.2. There exists $m_0 \ge 1$ such that

$$
\lim_{\delta\downarrow 0} \limsup_{t\to\infty} t\,\mathsf{P}\left[\frac{X_0}{b(t)} \le \delta,\ \sup_{j\ge m_0} \frac{X_j^{(b(t))}}{b(t)} > a\right] = 0 \qquad \text{for all } a > 0.
$$

This can be thought of as a moment condition, similar to (3.2.11): by Markov's inequality, it is sufficient that

$$\lim_{\delta\downarrow 0}\limsup_{t\to\infty} t\,\mathsf{E}\left[\left(\sup_{j\geq m_0}\frac{X_j^{(b(t))}}{b(t)}\right)^{\varepsilon}\mathbf{1}_{\{X_0\leq\delta b(t)\}}\right]=0$$

for some $\varepsilon>0$. Condition 3.2.2 implies a uniform bound on the α-th moment of the tail chain steps. Note that, if (3.2.9) holds, we can assume $m_0=1$ without loss of generality.

Proposition 3.2.4. *Let $\boldsymbol{X}\sim K\in D(G)$ be a Markov chain on $[0,\infty)$ with initial distribution F satisfying $1-F\in\mathrm{RV}_{-\alpha}$, whose extremal component is jointly regularly varying on $\mathbb{M}_+(\mathbb{E}^*)$ as in* (3.2.9) *(p. 53). Suppose that $\boldsymbol{X}$ satisfies Condition 3.2.2. Then*

$$\mathsf{E}\left(\sup_{j\geq 1}\xi(j)^{\alpha}\right)<\infty. \tag{3.2.14}$$

Remark. Under (3.2.14), we necessarily have $\mathsf{E}\,\xi_1^{\alpha}\leq 1$. This follows since

$$\sup_{j\geq 1}\left(\mathsf{E}\,\xi_1^{\alpha}\right)^j=\sup_{j\geq 1}\mathsf{E}\,\xi(j)^{\alpha}\leq\mathsf{E}\left(\sup_{j\geq 1}\xi(j)^{\alpha}\right)<\infty.$$

Recalling (3.2.10), this means that the marginal tails of the extremal component are no heavier that that of F.

Proof. Under (3.2.9),

$$\begin{aligned}&t\,\mathsf{P}\left[\sup_{m_0\leq j\leq r}X_j^{(b(t))}>b(t)\right]\\&\qquad\longrightarrow\int_{[0,\infty]}\nu_\alpha(dx)\,\mathsf{P}\left[\sup_{m_0\leq j\leq r}\xi(j)>x^{-1}\right]=\mathsf{E}\left(\sup_{m_0\leq j\leq r}\xi(j)^{\alpha}\right)\end{aligned}$$

by Lemma 3.2.1. Therefore,

$$\limsup_{t\to\infty}t\,\mathsf{P}\left[\sup_{j\geq m_0}X_j^{(b(t))}>b(t)\right]\geq\lim_{r\to\infty}\mathsf{E}\left(\sup_{m_0\leq j\leq r}\xi(j)^{\alpha}\right)=\mathsf{E}\left(\sup_{j\geq m_0}\xi(j)^{\alpha}\right).$$

Furthermore,

$$\begin{aligned}
&\limsup_{t\to\infty} t\,\mathsf{P}\left[\sup_{j\geq m_0} X_j^{(b(t))} > b(t)\right] \\
&\quad \leq \limsup_{t\to\infty} t\,\mathsf{P}\left[X_0 \leq \delta b(t), \sup_{j\geq m_0} X_j^{(b(t))} > b(t)\right] + \limsup_{t\to\infty} t\,\mathsf{P}\left[X_0 > \delta b(t)\right] < \infty
\end{aligned}$$

for some $\delta > 0$ by Condition 3.2.2, showing that $\mathsf{E}(\sup_{j\geq m_0} \xi(j)^\alpha) < \infty$. To verify (3.2.14), note that

$$\begin{aligned}
\mathsf{E}\left(\sup_{j\geq 1} \xi(j)^\alpha\right) &\leq \sum_{j=1}^{m_0-1} \mathsf{E}\,\xi(j)^\alpha + \mathsf{E}\left(\sup_{j\geq m_0} \xi(j)^\alpha\right) \\
&= \sum_{j=1}^{m_0-1} \left(\mathsf{E}\,\xi_1^\alpha\right)^j + \mathsf{E}\left(\sup_{j\geq m_0} \xi(j)^\alpha\right) < \infty. \qquad \square
\end{aligned}$$

Under the two conditions introduced above, we may derive the asymptotic distribution of the maximum of the extremal component of $\boldsymbol{X}$. Note that this is a maximum over a random number of observations.

Proposition 3.2.5. *Let $\boldsymbol{X} \sim K \in D(G)$ be a Markov chain on $[0,\infty)$ with initial distribution F satisfying $1-F \in \mathrm{RV}_{-\alpha}$, whose extremal component is jointly regularly varying on $\mathbb{M}_+(\mathbb{E}^*)$ as in (3.2.9). Suppose that $\boldsymbol{X}$ satisfies Conditions 3.2.1 and 3.2.2. Then*

$$t\,\mathsf{P}\left[\sup_{0\leq j<\tau(b(t))} \frac{X_j}{b(t)} \in \cdot\right] \xrightarrow{v} c\cdot \nu_\alpha(\cdot) \qquad \text{in } \mathbb{M}_+(0,\infty], \tag{3.2.15}$$

where

$$c = \mathsf{P}\left[\sup_{j\geq 1} \xi(j) \leq 1\right] + \mathsf{E}\left[\sup_{j\geq 1} \xi(j)^\alpha\, \mathbf{1}_{\{\sup_{j\geq 1}\xi(j)>1\}}\right].$$

Proof. For $m \geq 1$, write

$$t\,\mathsf{P}\left[\sup_{0\leq j<\tau(b(t))} \frac{X_j}{b(t)} > x\right] \leq t\,\mathsf{P}\left[\sup_{0\leq j<m} \frac{X_j^{(b(t))}}{b(t)} > x\right] + t\,\mathsf{P}\left[\sup_{j\geq m} \frac{X_j^{(b(t))}}{b(t)} > x\right].$$

One on hand,

$$\liminf_{t\to\infty} t\,\mathsf{P}\left[\sup_{0\le j<\tau(b(t))}\frac{X_j}{b(t)}>x\right]\ge\lim_{t\to\infty} t\,\mathsf{P}\left[\sup_{0\le j<m}\frac{X_j^{(b(t))}}{b(t)}>x\right]$$
$$=x^{-\alpha}+\int_{[0,x]}\nu_\alpha(du)\,\mathsf{P}_u\left[\sup_{1\le j<m}T_j>x\right]$$

by (3.2.9), from which

$$\liminf_{t\to\infty} t\,\mathsf{P}\left[\sup_{0\le j<\tau(b(t))}\frac{X_j}{b(t)}>x\right]\ge x^{-\alpha}+\int_{[0,x]}\nu_\alpha(du)\,\mathsf{P}_u\left[\sup_{j\ge1}T_j>x\right],\tag{3.2.16}$$

letting $m\to\infty$. On the other hand, for $\delta>0$ we have

$$t\,\mathsf{P}\left[\sup_{j\ge m}\frac{X_j^{(b(t))}}{b(t)}>x\right]\le t\,\mathsf{P}\left[\frac{X_0}{b(t)}>\delta,\ \sup_{j\ge m}\frac{X_j^{(b(t))}}{b(t)}>x\right]$$
$$+\,t\,\mathsf{P}\left[\frac{X_0}{b(t)}\le\delta,\ \sup_{j\ge m}\frac{X_j^{(b(t))}}{b(t)}>x\right].$$

Given $\varepsilon>0$, choose δ small enough that

$$\limsup_{t\to\infty} t\,\mathsf{P}\left[X_0\le\delta b(t),\ \sup_{j\ge m_0}X_j^{(b(t))}>b(t)x\right]<\varepsilon/2$$

by Condition 3.2.2. Next, choose $m_1\ge m_0$ large enough that

$$\limsup_{t\to\infty} t\,\mathsf{P}\left[X_0>\delta b(t),\ \sup_{j\ge m_1}X_j^{(b(t))}>b(t)x\right]<\varepsilon/2$$

by Condition 3.2.1. Therefore,

$$\limsup_{t\to\infty} t\,\mathsf{P}\left[\sup_{j\ge m}\frac{X_j^{(b(t))}}{b(t)}>x\right]<\varepsilon$$

for $m\ge m_1$, and so

$$\limsup_{t\to\infty} t\,\mathsf{P}\left[\sup_{0\le j<\tau(b(t))}\frac{X_j}{b(t)}>x\right]<\lim_{m\to\infty}\limsup_{t\to\infty} t\,\mathsf{P}\left[\sup_{0\le j<m}\frac{X_j^{(b(t))}}{b(t)}>x\right]+\varepsilon$$
$$=x^{-\alpha}+\int_{[0,x]}\nu_\alpha(du)\,\mathsf{P}_u\left[\sup_{j\ge1}T_j>x\right]+\varepsilon.$$

Combining this with (3.2.16), and applying the formula (3.2.3) (p. 49), completes the proof. □

3.3 Point Process Convergence for Markov Chains

We now turn to the question of deriving the limit of the exceedance point process N_n defined in (3.1.1), where $\boldsymbol{X} = (X_0, X_1, \dots)$ is a Markov chain on $[0, \infty)$ with transition kernel $K \in D(G)$. We write

$$N_n = \sum_{j=0}^{\infty} \epsilon_{\left(\frac{j}{n}, \frac{X_j}{b_n}\right)},$$

using the notation ϵ_x to denote the measure assigning unit mass at the point x, and we view N_n as a random element of $\mathbb{M}_p([0, \infty) \times (0, \infty])$, the space of Radon point measures on $[0, \infty) \times (0, \infty]$, endowed with the topology of vague convergence. Compact sets in this space are contained in rectangles of the form $[0, a] \times [b, \infty]$. For further details on convergence of random measures, see Resnick [76].

It is well-known that if $\boldsymbol{X}$ is positive recurrent, then it is a regenerative process (see Asmussen [2, Section VII.3]). This means that the sample path of $\boldsymbol{X}$ may be split into identically distributed "cycles" between visits to a certain set. The extremal properties of $\boldsymbol{X}$ may then be derived from those of the individual cycles. This approach has been developed for Markov chain exceedance times by Rootzén [81], as well as for queues by Asmussen [1]. We propose to use the tail chain approximation to describe the extremal behaviour of the regenerative cycles in detail via their extremal component.

3.3.1 Cycle Decomposition

Suppose $\boldsymbol{X}$ has a positive recurrent atom. This is a set A such that

$$K(y\,,\,\cdot) = H(\cdot) \quad \text{for all } y \in A \qquad \text{and} \qquad \mathsf{P}_y[\tau_A < \infty] = 1 \quad \text{for } y \geq 0, \qquad (3.3.1)$$

where H is a probability distribution on $[0,\infty)$ and $\tau_A = \inf\{n \geq 0 : X_n \in A\}$ is the first hitting time of A. Positive recurrence means that

$$\mathsf{E}_H\,\tau_A < \infty, \tag{3.3.2}$$

where E_H denotes expectation with respect to the initial distribution $X_0 \sim H$.

Under (3.3.1), the sample path of $\boldsymbol{X}$ splits into iid cycles between visits to A, as follows. Define the times $\{S_k\}$, $\{\tau_k^A\}$ recursively according to

$$\begin{aligned} \tau_0^A &= \tau_A, & S_0 &= \tau_0^A + 1; \\ \tau_k^A &= \inf\{n \geq 0 : X_{S_{k-1}+n} \in A\}, & S_k &= S_{k-1} + \tau_k^A + 1, \quad k \geq 1. \end{aligned} \tag{3.3.3}$$

Thus, the sequence $0 \leq S_0 - 1 < S_1 - 1 < S_2 - 1 < \cdots$ give the times at which $\boldsymbol{X}$ is in A, and $X_{S_k} \sim H$ for $k \geq 0$. The values $\tau_k^A \geq 0$ represent the number of steps $\boldsymbol{X}$ takes outside of A between visits to A. The cycles are random elements

$$C_0 = \big(X_0, X_1, \ldots, \underset{\substack{\in \\ A}}{X_{\tau_0^A}}\big) \quad \text{and} \quad C_k = \big(\underset{\substack{\wr \\ H}}{X_{S_{k-1}}}, \ldots, \underset{\substack{\in \\ A}}{X_{S_{k-1}+\tau_k^A}}\big), \quad k \geq 1$$

of the space of finite sequences $\mathcal{S} = \bigcup_{m=1}^{\infty} \mathbb{R}^m$. It is easy to see that $C_0, C_1, \ldots$ are independent, and $C_1, C_2, \ldots$ are identically distributed, by the strong Markov property. In particular,

$$\begin{aligned} \mathsf{P}\left[\{C_k\,;\,\tau_k^A\} \in \cdot\,\right] &= \mathsf{P}\left[\big\{(X_{S_{k-1}}, \ldots, X_{S_{k-1}+\tau_k^A})\,;\,\tau_k^A\big\} \in \cdot\,\right] \\ &= \mathsf{P}_H\left[\big\{(X_0, \ldots, X_{\tau_A})\,;\,\tau_A\big\} \in \cdot\,\right] \end{aligned} \tag{3.3.4}$$

for $k \geq 1$. Furthermore, $0 < S_0 < S_1 < S_2 < \cdots$ is a renewal process, with

$$q = \mathsf{E}(S_1 - S_0) = \mathsf{E}_H\,\tau_A + 1 < \infty \tag{3.3.5}$$

by (3.3.2). In terms of this cycle decomposition, we may now write

$$N_n = \sum_{j=0}^{\infty} \epsilon_{\left(\frac{j}{n},\, \frac{X_j}{b_n}\right)} = \sum_{j=0}^{S_0-1} \epsilon_{\left(\frac{j}{n},\, \frac{X_j}{b_n}\right)} + \sum_{k=0}^{\infty} \sum_{j=0}^{\tau_{k+1}^A} \epsilon_{\left(\frac{S_k+j}{n},\, \frac{X_{S_k+j}}{b_n}\right)} = \chi_n^0 + \chi_n^*. \tag{3.3.6}$$

First, we show that χ_n^0 is asymptotically negligible.

Proposition 3.3.1. *In* $\mathbb{M}_p([0,\infty)\times(0,\infty])$, *as* $n\to\infty$, $\vartheta_n+\chi_n^0 \Rightarrow \vartheta$ *iff* $\vartheta_n \Rightarrow \vartheta$.

Proof. This follows by an application of Slutsky's theorem [77, p. 55], provided $\mathsf{P}[d_v(\vartheta_n+\chi_n^0,\vartheta_n) > \gamma] \to 0$ for $\gamma > 0$. Here d_v denotes the vague metric. Let $f \in \mathcal{C}_K^+([0,\infty)\times(0,\infty])$ with support $[0,R]\times[M,\infty]$. It is sufficient to verify that $\mathsf{P}\left[|(\vartheta_n+\chi_n^0)(f)-\vartheta_n(f)| > \gamma\right] = \mathsf{P}[\chi_n^0(f) > \gamma] \to 0$. We have

$$\begin{aligned}\mathsf{P}\left[\chi_n^0(f) > \gamma\right] &= \mathsf{P}\left[\sum_{j=0}^{S_0-1} f\left(\frac{j}{n},\frac{X_j}{b_n}\right) > \gamma\right]\\ &\le \sum_{m=0}^{r}\mathsf{P}\left[\sum_{j=0}^{m} f\left(\frac{j}{n},\frac{X_j}{b_n}\right) > \gamma,\, \tau_A = m\right] + \mathsf{P}[\tau_A > r],\end{aligned}$$

and

$$\begin{aligned}\sum_{m=0}^{r}\mathsf{P}\left[\sum_{j=0}^{m} f\left(\frac{j}{n},\frac{X_j}{b_n}\right) > \gamma,\, \tau_A = m\right] &\le (r+1)\,\mathsf{P}\left[\sum_{j=0}^{r} f\left(\frac{j}{n},\frac{X_j}{b_n}\right) > \gamma\right]\\ &\le (r+1)\,\mathsf{P}\left[\sup_{0\le j\le r} X_j \ge b_n M\right] \longrightarrow 0\end{aligned}$$

as $n\to\infty$, recalling that $b_n\to\infty$. Choosing r to make $\mathsf{P}[\tau_A > r]$ arbitrarily small, the result follows. □

3.3.2 General Point Process Convergence Results

Proposition 3.3.1 shows that we can restrict consideration to the asymptotics of the point process χ_n^* in (3.3.6). A limit may be obtained using the tail chain approximation discussed in Section 3.2, provided that a cycle's extremal behaviour is adequately described by its extremal component.

We assume that $K \in D(G)$, with extremal boundary $y(t)$, and $1-H \in \mathrm{RV}_{-\alpha}$. Also, we require the atom A to be a bounded subset of the state space, i.e.,

$$\sup A < \infty; \tag{3.3.7}$$

this is indeed the case in most common scenarios. For $k = 1, 2, \ldots$, write

$$\tau_k(t) = \inf\big\{n \geq 0 : X_{S_{k-1}+n} \leq t y(t)\big\},$$

the index of the first element in the k-th cycle to cross below the extremal boundary. We think of $C_k(t) = \{X_{S_{k-1}+j} : j = 0, \ldots, \tau_k(t) - 1\}$ as the extremal component of the k-th cycle.

Note that, without loss of generality, the extremal component of a cycle is a subset of the complete cycle, since

$$\mathsf{P}[\tau(t) \leq \tau_A\,,\, t > 0] = 1. \tag{3.3.8}$$

Indeed, (3.3.7) implies that $A \subset [0, c]$ some c. Define $\tau_c = \inf\{n \geq 0 : X_n \leq c\}$. Clearly $\mathsf{P}[\tau_c \leq \tau_A] = 1$; we claim further that $\mathsf{P}[\tau(t) \leq \tau_c,\, t > 0] = 1$. If $y(t) \geq c/t$ for $t > 0$, then this follows directly. Otherwise, it is easy to see that $\tilde{y}(t) = y(t) \vee c/t$ is also an extremal boundary for K, and the corresponding downcrossing time satisfies $\mathsf{P}[\tilde{\tau}(t) \leq \tau_c,\, t > 0] = 1$.

Therefore,

$$\begin{aligned}\mathsf{P}\left[\big\{(X_{S_{k-1}}, \ldots, X_{S_{k-1}+\tau_k(t)-1})\,;\, \tau_k(t), \tau_k^A\big\} \in \cdot\,\right] \\ = \mathsf{P}_H\left[\big\{(X_0, \ldots, X_{\tau(t)-1})\,;\, \tau(t), \tau_A\big\} \in \cdot\,\right]\end{aligned} \tag{3.3.9}$$

for $k \geq 1$, and the $\big\{\{C_k(t);\, \tau_k(t), \tau_k^A\} : k \geq 1\big\}$ are independent, since each is a function of $\{C_k;\, \tau_k^A\}$. In light of these facts, we approximate χ_n^* by a point process whose observations consist of iid copies of the extremal component of the Markov chain $\boldsymbol{X}'$ with transition kernel K and initial distribution $X_0' \sim H$.

At this point we must introduce some notation. Let $\{\boldsymbol{X}_k : k = 0, 1, \ldots\}$ be iid copies of the Markov chain $\boldsymbol{X}' \sim K$ with initial distribution H. Define

$$\tilde{\tau}_{k+1}(t) = \inf\big\{j \geq 0 : X_{kj} \leq t y(t)\big\}, \qquad k = 0, 1, \ldots,$$

and form the respective extremal components

$$\big\{\boldsymbol{X}_k^{(t)} = \{X_{kj}\cdot \mathbf{1}_{\{j<\tilde\tau_{k+1}(t)\}},\, j\geq 0\} : k=0,1,\dots\big\}.$$

Then $\big(\boldsymbol{X}_k^{(t)},\, \tilde\tau_{k+1}(t)\big) \stackrel{d}{=} \big(\boldsymbol{X}'^{(t)},\, \tau'(t)\big)$ for all k, with the tilde differentiating the iid times $\tilde\tau_k(t)$ from the cycle times $\tau_k(t)$.

Next, let $\{\boldsymbol{\xi}_k = (\xi_k(0), \xi_k(1), \dots) : k = 0,1,\dots\}$ be iid copies of the process $\boldsymbol{\xi} = (\xi(0), \xi(1), \dots)$, recalling the notation $\xi(n) = \prod_{j=1}^n \xi_j$, with $\xi(0) = 1$. Additionally, put $\tau^*_{k+1} = \inf\{j\geq 0 : \xi_k(j) = 0\}$. We adopt the convention $\inf\varnothing = \infty$, so that $\mathsf{P}[\tau^*_{k+1} = \infty] = 1$ when $G(\{0\}) = 0$.

Finally, set

$$\zeta = \sum_{k=0}^{\infty} \epsilon_{(t_k, i_k)} \sim \mathrm{PRM}(\mathbb{LEB}\times\nu_\alpha),$$

a Poisson random measure on $\mathbb{M}_p([0,\infty)\times(0,\infty])$, independent of the $\{\boldsymbol{\xi}_k\}$, with mean measure given by the product of Lebesgue measure on the time axis $[0,\infty)$ with Pareto measure ν_α on the observation axis $(0,\infty]$. Recall $\nu_\alpha(x,\infty] = x^{-\alpha}$, and α is the tail index of H.

Consequently, the point process consisting of the observations $\boldsymbol{X}_k^{(b_n)}$, spaced out in time according to the renewal times $\{S_k\}$, converges to a compound Poisson process, compounded according to the $\{\boldsymbol{\xi}_k\}$.

Theorem 3.3.1. *Suppose $\boldsymbol{X}'$ is a Markov chain on $[0,\infty)$ with transition kernel $K \in D(G)$ and initial distribution $X'_0 \sim H$ satisfying $tH(b(t)\cdot) \xrightarrow{v} \nu_\alpha$ in $\mathbb{M}_+(0,\infty]$, where $b(t)\to\infty$.*

(a) If $\boldsymbol{X}'$ satisfies Condition 3.2.1 (p. 56), then, given $\delta > 0$,

$$\eta_n = \sum_{k=0}^{\infty}\sum_{j=0}^{\tilde\tau_{k+1}(b_n)-1} \epsilon_{\left(\frac{S_k+j}{n},\, \frac{X_{kj}}{b_n}\right)} \mathbf{1}_{\{X_{k0}\geq \delta b_n\}} \Rightarrow \sum_{k=0}^{\infty}\sum_{j=0}^{\tau^*_{k+1}-1} \epsilon_{(qt_k,\, i_k\xi_k(j))}\mathbf{1}_{\{i_k\geq\delta\}} = \eta$$

in $\mathbb{M}_p([0,\infty)\times(0,\infty])$ *as* $n\to\infty$.

(b) Suppose additionally that $\boldsymbol{X}'^{(t)}$ *is jointly regularly varying, i.e., the convergence* (3.2.9) *(p. 53) takes place. If* $\boldsymbol{X}'$ *satisfies both Conditions 3.2.1 and 3.2.2, then*

$$\eta_n^* = \sum_{k=0}^{\infty} \sum_{j=0}^{\tilde{\tau}_{k+1}(b_n)-1} \epsilon_{\left(\frac{S_k+j}{n},\, \frac{X_{kj}}{b_n}\right)} \Rightarrow \sum_{k=0}^{\infty} \sum_{j=0}^{\tau_{k+1}^*-1} \epsilon_{(qt_k,\, i_k\xi_k(j))} = \eta^*$$

in $\mathbb{M}_p([0,\infty)\times(0,\infty])$ *as* $n\to\infty$.

Here $\{(\boldsymbol{X}_k,\, \tilde{\tau}_{k+1}(t))\}$ *and* $\{(\boldsymbol{\xi}_k,\, \tau_{k+1}^*)\}$ *are as above;* $b_n = b(n)$*;* $\{S_k\}$ *is the renewal process defined in* (3.3.3)*, with mean interarrival time* q *given by* (3.3.5)*; and* (t_k, i_k) *are the points of* ζ.

The proof is given in Section 3.4 (p. 78). In keeping with the discussion in Section 3.2, we present two convergences, depending on the strength of our conditions. The weaker assumptions of H regularly varying together with the stability provided under Condition 3.2.1 yield a conditional result, where the point process records those cycles which start from an exceedance. The unconditional result requires joint regular variation, as well as the moment restriction given by Condition 3.2.2.

Note that the points of the limit process are arranged in stacks above common timepoints qt_k. The heights of the points in each stack are determined by an independent run of the tail chain starting from i_k. If $G(\{0\}) > 0$, then the τ_k^* are iid Geometric random variables with parameter $G(\{0\})$, so the stacks all have finite length. If $G(\{0\}) = 0$, then $\mathsf{P}[\tau_k^* = \infty] = 1$ for each k. In this case, Condition 3.2.1 ensures that η^* is Radon by forcing the tail chain to drift towards 0 as in (3.2.13). The process η retains only those stacks of η^* whose initial value exceeds the threshold δ. Because there are an infinite number of i_k in any neighbourhood of 0, dispensing with the restriction in δ requires that not too many of the $\xi_k(j)$ are large. This translates to the constraint $\mathsf{E}\,\xi_1^\alpha \le 1$, which follows from Condition 3.2.2.

Returning to the original chain $\boldsymbol{X}$, we can approximate χ_n^* by η_n^*, provided that the extremal component is sufficient to describe the extremal behaviour within each cycle. What happens between the end of the extremal component and the end of the cycle will not be captured by the tail chain. Therefore, we require that these observations not influence the asymptotics. We formulate this as

Condition 3.3.1.

$$\lim_{t\to\infty} \mathsf{P}\left[\sup_{\tau(b(t))<j<\tau_A} \frac{X_j}{b(t)} > a \,\middle|\, X_0 > \delta b(t) \right] = 0 \qquad \text{for all } \ a, \delta > 0$$

in the conditional case, and

Condition 3.3.2.

$$\lim_{t\to\infty} t\,\mathsf{P}\left[\sup_{\tau(b(t))<j<\tau_A} \frac{X_j}{b(t)} > a \right] = 0 \qquad \text{for all } \ a > 0$$

in the unconditional one. Under the latter condition, the point process N_n converges to the limit η^*, and the distribution of the cycle maximum has a regularly varying tail. In view of (3.3.4) and (3.3.9), assumptions on the distribution of the cycles $\{C_k\}$ will be phrased in terms of P_H.

Theorem 3.3.2. *Let $\boldsymbol{X}$ be a Markov chain on $[0,\infty)$ with transition kernel $K \in D(G)$. Suppose that K has a positive recurrent bounded atom in the sense of* (3.3.1), (3.3.2), *and* (3.3.7). *Define the renewal process $\{S_k\}$ with mean interarrival time q as in* (3.3.3) *and* (3.3.5). *Assume further that $tH(b(t)\cdot) \xrightarrow{v} \nu_\alpha$ in $\mathbb{M}_+(0,\infty]$, where $b(t) \to \infty$.*

(a) *If $\boldsymbol{X}$ satisfies Conditions 3.2.1 and 3.3.1 with respect to P_H, i.e., when $X_0 \sim H$, then, given $\delta > 0$,*

$$\tilde{N}_n = \sum_{j=0}^{\infty} \epsilon_{\left(\frac{j}{n}, \frac{X_j}{b_n}\right)} \left\{ \mathbf{1}_{\{j<S_0\}} + \sum_{k=1}^{\infty} \mathbf{1}_{\{S_{k-1} \le j < S_k\}} \mathbf{1}_{\{X_{S_{k-1}} \ge \delta b_n\}} \right\} \Rightarrow \eta \quad (3.3.10)$$

in $\mathbb{M}_p([0,\infty) \times (0,\infty])$, as $n \to \infty$.

(b) *Suppose additionally that $\boldsymbol{X}^{(t)}$ is jointly regularly varying when $X_0 \sim H$, i.e., the convergence (3.2.9) takes place. If $\boldsymbol{X}$ satisfies Conditions 3.2.1 and 3.2.2, as well as Condition 3.3.2, with respect to P_H, then*

$$N_n \Rightarrow \eta^* \quad (3.3.11)$$

in $\mathbb{M}_p([0,\infty) \times (0,\infty])$, as $n \to \infty$. Furthermore, the distribution of the cycle maximum has a regularly varying tail:

$$t\,\mathsf{P}_H\left[\sup_{0 \le j < \tau_A} \frac{X_j}{b(t)} \in \cdot\right] \xrightarrow{v} c \cdot \nu_\alpha(\cdot) \qquad \text{in } \mathbb{M}_+(0,\infty], \quad (3.3.12)$$

where

$$c = \mathsf{P}\left[\sup_{j \ge 1} \xi(j) \le 1\right] + \mathsf{E}\left[\sup_{j \ge 1} \xi(j)^\alpha \mathbf{1}_{\{\sup_{j\ge1} \xi(j) > 1\}}\right]. \quad (3.3.13)$$

Proof. (a) First, note that $\tilde{N}_n = \chi_n^0 + \chi_n$, where

$$\chi_n = \sum_{k=0}^{\infty} \sum_{j=0}^{\tau_{k+1}^A} \epsilon_{\left(\frac{S_k+j}{n}, \frac{X_{S_k+j}}{b_n}\right)} \mathbf{1}_{\{X_{S_k} \ge \delta b_n\}}\,.$$

Hence, by Proposition 3.3.1 it remains to show that $\chi_n \Rightarrow \eta$. Split χ_n according to the times $\{\tau_k(t)\}$:

$$\begin{aligned}\chi_n &= \sum_{k=0}^{\infty} \sum_{j=0}^{\tau_{k+1}(b_n)-1} \epsilon_{\left(\frac{S_k+j}{n}, \frac{X_{S_k+j}}{b_n}\right)} \mathbf{1}_{\{X_{S_k} \ge \delta b_n\}} + \sum_{k=0}^{\infty} \sum_{j=\tau_{k+1}(b_n)}^{\tau_{k+1}^A} \epsilon_{\left(\frac{S_k+j}{n}, \frac{X_{S_k+j}}{b_n}\right)} \mathbf{1}_{\{X_{S_k} \ge \delta b_n\}} \\ &= \chi_n' + \chi_n''.\end{aligned}$$

The equality holds on the set $\{\tau_k(b_n) \le \tau_k^A \,;\, n \ge 1, k \ge 1\}$, which has probability 1 by (3.3.8). Because of (3.3.9) and the independence of the $(C_k(t), \tau_k(t))$, we have $\chi_n' \overset{d}{=} \eta_n$ for each n, and $\eta_n \Rightarrow \eta$ by Theorem 3.3.1 (a). Therefore, by Slutsky's theorem, the result will follow once we show that $d_v(\chi_n, \chi_n') \xrightarrow{P} 0$, where d_v is the vague metric. This entails

$$\mathsf{P}\left[\chi_n''(f) > \gamma\right] = \mathsf{P}\left[\sum_{k=0}^{\infty} \sum_{j=\tau_{k+1}(b_n)}^{\tau_{k+1}^A} f\left(\frac{S_k + j}{n}, \frac{X_{S_k+j}}{b_n}\right) \mathbf{1}_{\{X_{S_k} \ge \delta b_n\}} > \gamma\right] \longrightarrow 0$$

for any $f \in \mathcal{C}_K^+([0,\infty) \times (0,\infty])$. Let f have support $[0,R] \times [M,\infty]$. The previous probability is bounded by

$$\mathsf{P}\left[\sum_{k=0}^{2Rn-1} \sum_{j=\tau_{k+1}(b_n)}^{\tau_{k+1}^A} f\left(\frac{S_k + j}{n}, \frac{X_{S_k+j}}{b_n}\right) \mathbf{1}_{\{X_{S_k} \ge \delta b_n\}} > 0\right] + \mathsf{P}\left[\sum_{k=2Rn}^{\infty} \sum_{j=\tau_{k+1}(b_n)}^{\tau_{k+1}^A} f\left(\frac{S_k + j}{n}, \frac{X_{S_k+j}}{b_n}\right) \mathbf{1}_{\{X_{S_k} \ge \delta b_n\}} > 0\right].$$

The second term is at most $\mathsf{P}[S_{2Rn}/n \le R] = \mathsf{P}[S_{2Rn}/2Rn \le 1/2] \to 0$ as $n \to \infty$, since $S_n/n \to q$ a.s., and $q \ge 1$ by (3.3.5). The first term is bounded by

$$\mathsf{P}\left[\bigcup_{k=0}^{2Rn-1} \left(\left\{\frac{X_{S_k}}{b_n} \ge \delta\right\} \cap \bigcup_{j=\tau_{k+1}(b_n)}^{\tau_{k+1}^A} \left\{\frac{X_{S_k+j}}{b_n} \ge M\right\}\right)\right] \le 2Rn\, \mathsf{P}_H\left[\frac{X_0}{b_n} \ge \delta, \sup_{\tau(b_n) < j < \tau_A} \frac{X_j}{b_n} \ge M\right],$$

which vanishes as $n \to \infty$ by Condition 3.3.1.

(b) Recalling the decomposition (3.3.6) (p. 63), by Proposition 3.3.1 it is sufficient to show that $\chi_n^* \Rightarrow \eta^*$. This follows by a similar argument as in part (a). Write

$$\chi_n^* = \sum_{k=0}^{\infty} \sum_{j=0}^{\tau_{k+1}(b_n)-1} \epsilon_{\left(\frac{S_k+j}{n}, \frac{X_{S_k+j}}{b_n}\right)} + \sum_{k=0}^{\infty} \sum_{j=\tau_{k+1}(b_n)}^{\tau_{k+1}^A} \epsilon_{\left(\frac{S_k+j}{n}, \frac{X_{S_k+j}}{b_n}\right)} = \chi_n^{*\prime} + \chi_n^{*\prime\prime}.$$

Then $\chi_n^{*\prime} \stackrel{d}{=} \eta_n^* \Rightarrow \eta^*$ by Theorem 3.3.1 (b), and Condition 3.3.2 implies that $d_v(\chi_n^*, \chi_n^{*\prime}) \xrightarrow{P} 0$.

Next, we show (3.3.12). In light of (3.3.8) (p. 65), we have

$$\begin{aligned} 0 &\le t\,\mathsf{P}_H\left[\sup_{0\le j<\tau_A} \frac{X_j}{b(t)} > x\right] - t\,\mathsf{P}_H\left[\sup_{0\le j<\tau(b(t))} \frac{X_j}{b(t)} > x\right] \\ &\le t\,\mathsf{P}_H\left[\sup_{\tau(b(t))<j<\tau_A} \frac{X_j}{b(t)} > x\right] \longrightarrow 0 \end{aligned}$$

under Condition 3.3.2. Recalling that

$$t\,\mathsf{P}_H\left[\sup_{0\le j<\tau(b(t))} \frac{X_j}{b(t)} > x\right] \longrightarrow cx^{-\alpha}$$

as $t \to \infty$ by Proposition 3.2.5 (p. 60), where c is given by (3.3.13), the result follows. □

Writing $M_n = \max_{0\le j\le n} X_j$, Rootzén shows [81, Theorem 3.2] that (3.3.12) further implies

$$\mathsf{P}[M_n \le b_n x] \longrightarrow \exp(-cq^{-1}x^{-\alpha}), \qquad x > 0,$$

where c is given by (3.3.13), and q is the mean interarrival time (3.3.5). Hence, in the stationary case, $\theta = c/q$ is the extremal index of the process $\boldsymbol{X}$ (see Leadbetter and Rootzén [56, Section 2.2] for details). On the other hand, for stationary regularly varying Markov chains with $K \in D(G)$ satisfying a condition analogous to Condition 3.2.1, it is known that

$$\begin{aligned} \theta &= \mathsf{P}\left[\sup_{j\ge 1} Y\xi(j) \le 1\right] \\ &= \mathsf{P}\left[\sup_{j\ge 1} \xi(j) \le 1\right] - \mathsf{E}\left[\sup_{j\ge 1} \xi(j)^\alpha\, \mathbf{1}_{\{\sup_{j\ge1}\xi(j)\le 1\}}\right] \\ &= c - \mathsf{E}\left(\sup_{j\ge 1} \xi(j)^\alpha\right), \end{aligned}$$

where $Y \sim \text{Pareto}(\alpha)$ supported on $[1, \infty)$, independent of $\{\xi(j)\}$ (see Basrak and Segers [8, Remark 4.7]). Hence, for a stationary Markov chain $\boldsymbol{X}$ satisfying the assumptions of Theorem 3.3.2 (b), the extremal index is given by

$$\theta = \frac{1}{q-1}\,\mathsf{E}\left(\sup_{j\geq 1} \xi(j)^{\alpha}\right) = \frac{\mathsf{E}\left(\sup_{j\geq 1} \xi(j)^{\alpha}\right)}{\mathsf{E}_H\,\tau_A}.$$

We now consider simplifications of the above conditions, depending on whether $G(\{0\}) > 0$ or $G(\{0\}) = 0$.

3.3.3 Cases where $G(\{0\}) = 0$

If $G(\{0\}) = 0$, we can replace $\boldsymbol{X}^{(b(t))}$ with $\boldsymbol{X}$ in the finite-dimensional convergence (3.2.2) (p. 48) when F_{X_0} has a regularly varying tail, meaning that the tail chain approximation completely describes the extremes of the chain $\boldsymbol{X}$ in finite dimensions. However, $G(\{0\}) = 0$ implies that

$$\mathsf{P}_t[m < \tau(t) \leq \tau_A] \longrightarrow 1 \qquad \text{as } t \to \infty \tag{3.3.14}$$

for any m, meaning that, as the initial observation becomes more extreme, it will take longer for $\boldsymbol{X}$ to return to A to complete the cycle. Hence, for Condition 3.2.1 to hold, we need to ensure that $\boldsymbol{X}$ will eventually drift away from the extreme states:

Condition 3.3.3.

$$\lim_{m\to\infty} \limsup_{t\to\infty} \mathsf{P}_t\left[\sup_{m\leq j<\tau_A} X_j > ta\right] = 0 \qquad \text{for all } a > 0.$$

This is essentially a condition on the transition kernel K. In particular, it was shown in Section 2.2.3 (p. 15) that any transition kernel $K \in D(G)$ has a corresponding update function of the form

$$\psi(x,(Z,W)) = Zx + \phi(x,W), \tag{3.3.15}$$

where $Z \sim G$ and $t^{-1}\phi(t,w) \to 0$ for $w \in C$ with $\mathsf{P}[W \in C] = 1$. Take $V_r = (Z_r, W_r)$, iid copies of $V = (Z,W)$, and write $\boldsymbol{V}_r = (V_1,\dots,V_r)$. For $r \geq 1$ let $\psi^r(x,\boldsymbol{V}_r)$ denote the r-step update function, i.e., $K^r(x,B) = \mathsf{P}[\psi^r(x,\boldsymbol{V}_r) \in B]$, and $\psi^0(x) = x$. It is easy to see that

$$\begin{aligned}
&\psi^r(x,\boldsymbol{V}_r)\\
&\quad= \Big(\prod_{j=1}^r Z_j\Big)x + \sum_{\ell=1}^{r-1}\Big(\prod_{j=\ell+1}^r Z_j\Big)\phi\big(\psi^{\ell-1}(x,\boldsymbol{V}_{\ell-1}),W_\ell\big) + \phi\big(\psi^{r-1}(x,\boldsymbol{V}_{r-1}),W_r\big).
\end{aligned}$$

This suggests that, in addition to the requirement that $\prod_{j=1}^m Z_j \to 0$ a.s., Condition 3.3.3 translates to an asymptotic boundedness condition on the function $\phi(\cdot,W)$.

Proposition 3.3.2. *Suppose $\boldsymbol{X} \sim K \in D(G)$ with $G(\{0\}) = 0$ has a positive recurrent, bounded atom A, and $1-H \in \mathrm{RV}_{-\alpha}$. Then, under Condition 3.3.3, both Conditions 3.2.1 and 3.3.1 hold with respect to P_H. Consequently, the convergence* (3.3.10) *(p. 69) takes place.*

***Proof*.** First, we show that

$$\lim_{m\to\infty}\limsup_{t\to\infty} t\,\mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta, \sup_{m\leq j<\tau_A}\frac{X_j}{b(t)} > a\right] = 0 \qquad \text{for all } a,\delta > 0. \tag{3.3.16}$$

Indeed, for $c > \delta$, we have

$$\begin{aligned}
&t\,\mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta, \sup_{m\leq j<\tau_A}\frac{X_j}{b(t)} > a\right]\\
&\qquad\leq \int_{[\delta,c]} t\,\mathsf{P}_H\left[\frac{X_0}{b(t)} \in du\right]\mathsf{P}_{b(t)u}\left[\sup_{m\leq j<\tau_A}\frac{X_j}{b(t)} > a\right] + t\,\mathsf{P}_H[X_0 > b(t)c].
\end{aligned}$$

Furthermore, for $\delta \leq u \leq c$,

$$\begin{aligned}
\mathsf{P}_{b(t)u}\left[\sup_{m\leq j<\tau_A}\frac{X_j}{b(t)} > a\right] &\leq \mathsf{P}_{b(t)u}\left[\sup_{m\leq j<\tau_A}\frac{X_j}{b(t)u} > \frac{a}{c}\right]\\
&\leq \sup_{s\geq b(t)\delta}\mathsf{P}_s\left[\sup_{m\leq j<\tau_A}\frac{X_j}{s} > \frac{a}{c}\right].
\end{aligned}$$

Hence, by Condition 3.3.3,

$$\lim_{m\to\infty}\limsup_{t\to\infty} t\,\mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta,\ \sup_{m\le j<\tau_A}\frac{X_j}{b(t)} > a\right] \le \nu_\alpha[\delta,c]\cdot 0 + \nu_\alpha(c,\infty] = c^{-\alpha}.$$

Letting $c \to \infty$ establishes (3.3.16). As (3.3.8) implies that

$$\sup_{m\le j<\tau(b(t))} X_j \le \sup_{m\le j<\tau_A} X_j,$$

Condition 3.2.1 follows. To verify Condition 3.3.1, argue that

$$\begin{aligned}\mathsf{P}_H&\left[\frac{X_0}{b(t)} > \delta,\ \sup_{\tau(b(t))<j<\tau_A}\frac{X_j}{b(t)} > a\right]\\ &\le \mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta,\ \tau(b(t)) \le m-1\right] + \mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta,\ \sup_{m\le j<\tau_A}\frac{X_j}{b(t)} > a\right],\end{aligned}$$

of which the first term vanishes as $t \to \infty$ by Proposition 3.2.1 (d) (p. 48)—this is a consequence of (3.3.14). Appealing to (3.3.16) completes the proof as $m \to \infty$ subsequently. □

3.3.4 Cases where $G(\{0\}) > 0$

Here

$$\mathsf{P}_{tu}[\tau(t) = m] \longrightarrow \mathsf{P}[\tau^* = m], \qquad m \ge 1,$$

where τ^* is a Geometric random variable with parameter $G(\{0\})$. Hence, the tail chain terminates after a finite number of steps. If either $y_0(t) \equiv 0$ is an extremal boundary, or K satisfies the regularity condition (2.4.3) (p. 33), the regular variation (3.2.2) holds for $\boldsymbol{X}$ with respect to P_H, and Condition 3.2.1 follows directly. It is easy to show that the regularity condition extends to any finite number of steps. However, unless $y_0(t) \equiv 0$ is an extremal boundary, we need the regularity condition to hold uniformly over the whole cycle of random length τ_A. This prevents $\boldsymbol{X}$ from returning to an extreme state within the same cycle, after crossing below the extremal boundary. Note that, even if $y_0(t)$ is an extremal boundary for K, we are using an extremal boundary $y(t)$ chosen to satisfy (3.3.8).

Condition 3.3.4.

$$\lim_{t\to\infty} \mathsf{P}_{tu_t}\left[\sup_{1\le j<\tau_A} X_j > ta\right] = 0 \qquad \text{whenever } u_t = u(t) \to 0, \qquad a > 0.$$

Recalling the update function form (3.3.15), we found the regularity condition to hold if the function ϕ is bounded near 0 (Proposition 2.4.1, p. 33). We can think of Condition 3.3.4 as a stronger boundedness restriction. Alternatively, assuming that K satisfies the regularity condition, Condition 3.3.4 may be viewed as a restriction on τ_A, since it is sufficient that

$$\lim_{m\to\infty} \limsup_{t\to\infty} \mathsf{P}_{tu_t}[\tau_A > m] = 0 \qquad \text{whenever } u_t = u(t) \to 0.$$

Proposition 3.3.3. *Suppose $\boldsymbol{X} \sim K \in D(G)$ with $G(\{0\}) > 0$ has a positive recurrent, bounded atom A, and $1 - H \in \mathrm{RV}_{-\alpha}$. Then $\boldsymbol{X}$ satisfies Condition 3.2.1 with respect to P_H. Moreover, if either $y_0(t) \equiv 0$ is an extremal boundary for K, or else if Condition 3.3.4 holds, then Condition 3.3.1 holds with respect to P_H. Consequently, the convergence* (3.3.10) *(p. 69) takes place.*

Proof. First, note that

$$t\,\mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta,\, \sup_{j\ge m} \frac{X_j^{(b(t))}}{b(t)} > a\right] \tag{3.3.17}$$
$$\le t\,\mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta,\, \tau(b(t)) > m\right] \longrightarrow \delta^{-\alpha}\big(1 - G(\{0\})\big)^m$$

as $t \to \infty$ by Proposition 3.2.1 (d) (p. 48). This quantity vanishes as $m \to \infty$, establishing Condition 3.2.1. Next, consider the case where $y_0(t) \equiv 0$ is an extremal boundary, and write $\tau_0 = \inf\{n \geq 0 : X_n = 0\}$. For any m,

$$t\,\mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta,\ \sup_{\tau(b(t))<j<\tau_A} \frac{X_j}{b(t)} > a\right]$$
$$\leq \sum_{r=1}^{m} t\,\mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta,\ \tau(b(t)) = r,\ \tau_0 > r\right] + t\,\mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta,\ \tau(b(t)) > m\right].$$

For the first term, we have

$$t\,\mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta, \tau(b(t)) = r,\ \tau_0 > r\right]$$
$$\leq t\,\mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta,\ \tau_0 > r\right] - t\,\mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta,\ \tau(b(t)) > r\right]$$
$$\longrightarrow \delta^{-\alpha}\big(1 - G(\{0\})\big)^r - \delta^{-\alpha}\big(1 - G(\{0\})\big)^r = 0$$

since both y and y_0 are extremal boundaries. The second term is handled as in (3.3.17). Suppose now that Condition 3.3.4 holds. For any m, we have

$$t\,\mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta,\ \sup_{\tau(b(t))<j<\tau_A} \frac{X_j}{b(t)} > a\right]$$
$$\leq \sum_{r=1}^{m} t\,\mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta,\ \sup_{r<j<\tau_A} \frac{X_j}{b(t)} > a,\ \tau(b(t)) = r\right]$$
$$+ t\,\mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta,\ \tau(b(t)) > m\right],$$

and

$$t\,\mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta,\ \sup_{r<j<\tau_A} \frac{X_j}{b(t)} > a,\ \tau(b(t)) = r\right] \tag{3.3.18}$$
$$= \int_{[\delta,\infty]\times(y(t),\infty]^{m-1}\times[0,\infty]} t\,\mathsf{P}_H\left[\left(\frac{X_0}{b(t)}, \frac{\boldsymbol{X}_m}{b(t)}\right) \in d(x_0, \boldsymbol{x}_m)\right] h_t(x_m),$$

where

$$h_t(x) = \mathbf{1}_{\{[0,y(t)]\}}(x)\,\mathsf{P}_{b(t)x}\left[\sup_{1\leq j<\tau_A} \frac{X_j}{b(t)} > a\right].$$

We claim that $h_t(u_t) \to 0$ whenever $u_t \to u \geq 0$. Indeed, if $u > 0$, then $h_t(u_t) = 0$ for large t since $y(t) < u$. Otherwise, $u_t \to 0$, and $h_t(u_t) \to 0$ by Condition 3.3.4. Therefore, by combining Lemmas A.0.2 and A.0.4 (p. 133) with Theorem 2.3.2 (p. 26), we find that the integral in (3.3.18) converges to 0. Applying (3.3.17) completes the proof. □

3.3.5 Jointly Regularly Varying Markov Chains

If it is the case that the finite-dimensional distributions of $\boldsymbol{X}$ are jointly regularly varying, or equivalently $t\,\mathsf{P}[X_j > b(t)x] \to (\mathsf{E}\,\xi_1^\alpha)^j x^{-\alpha}$ for $j \geq 0$ (Theorem 3.2.1, p. 53, with $\boldsymbol{X}$ replacing $\boldsymbol{X}^{(b(t))}$), then we obtain a point process limit for the complete chain $\boldsymbol{X}$ under a moment condition analogous to Condition 3.2.2:

Condition 3.3.5. There exists $m_0' \geq 1$ such that

$$\lim_{\delta\downarrow 0} \limsup_{t\to\infty} t\,\mathsf{P}\left[\frac{X_0}{b(t)} \leq \delta, \sup_{m_0' \leq j < \tau_A} \frac{X_j}{b(t)} > a\right] = 0 \qquad \text{for all } a > 0.$$

Proposition 3.3.4. *Suppose $\boldsymbol{X} \sim K \in D(G)$ has a positive recurrent, bounded atom A, and $1 - H \in \mathrm{RV}_{-\alpha}$. Assume further that, with respect to P_H, $\boldsymbol{X}$ is regularly varying in the sense of* (3.2.9) *(p. 53), with $\boldsymbol{X}$ replacing $\boldsymbol{X}^{(b(t))}$, and satisfies Condition 3.3.1. Under Condition 3.3.5, both Conditions 3.2.2 and 3.3.2 hold with respect to P_H.*

Proof. First, recalling $\sup_{m\leq j<\tau(b(t))} X_j \leq \sup_{m \leq j < \tau_A} X_j$ yields (3.2.2). Next, given $\delta > 0$, write

$$\begin{aligned}
&\mathsf{P}_H\left[\sup_{\tau(b(t))<j<\tau_A} \frac{X_j}{b(t)} > a\right] \\
&\qquad \leq \mathsf{P}_H\left[\frac{X_0}{b(t)} > \delta, \sup_{\tau(b(t))<j<\tau_A} \frac{X_j}{b(t)} > a\right] + \mathsf{P}_H\left[\frac{X_0}{b(t)} \leq \delta, \sup_{1\leq j<\tau_A} \frac{X_j}{b(t)} > a\right].
\end{aligned}$$

By Condition 3.3.1,

$$\begin{aligned}
\limsup_{t\to\infty} \mathsf{P}_H\left[\sup_{\tau(b(t))<j<\tau_A} \frac{X_j}{b(t)} > a\right] &\leq \limsup_{t\to\infty} \mathsf{P}_H\left[\frac{X_0}{b(t)} \leq \delta, \sup_{1\leq j<\tau_A} \frac{X_j}{b(t)} > a\right] \\
&\leq \limsup_{t\to\infty} \mathsf{P}_H\left[\frac{X_0}{b(t)} \leq \delta, \sup_{1\leq j<m_0'} \frac{X_j}{b(t)} > a\right] \\
&\qquad + \limsup_{t\to\infty} \mathsf{P}_H\left[\frac{X_0}{b(t)} \leq \delta, \sup_{m_0'\leq j<\tau_A} \frac{X_j}{b(t)} > a\right] \\
&= \mu^*([0,\delta] \times [\mathbf{0}, \boldsymbol{a}]^{\mathrm{c}}) + \limsup_{t\to\infty} \mathsf{P}_H\left[\frac{X_0}{b(t)} \leq \delta, \sup_{m_0'\leq j<\tau_A} \frac{X_j}{b(t)} > a\right]
\end{aligned}$$

where $\boldsymbol{a} = (a, \ldots, a)$. Letting $\delta \downarrow 0$, the first term vanishes by (3.2.9), and the second is taken care of by Condition 3.3.5. □

We now rephrase Theorem 3.3.2 in terms of our new conditions.

Theorem 3.3.3. *Let $\boldsymbol{X}$ be a Markov chain on $[0, \infty)$ with transition kernel $K \in D(G)$. Suppose that K has a positive recurrent, bounded atom in the sense of (3.3.1), (3.3.2), and (3.3.7), where H satisfies $tH(b(t)\cdot) \xrightarrow{v} \nu_\alpha$ in $\mathbb{M}_+(0, \infty]$. Assume further that*

$$t\,\mathsf{P}_H\left[\left(\frac{X_0}{b(t)}, \ldots, \frac{X_m}{b(t)}\right) \in (dx_0, d\boldsymbol{x}_m)\right] \xrightarrow{v} \nu_\alpha(dx_0)\,\mathsf{P}_{x_0}\left[(T_1, \ldots, T_m) \in d\boldsymbol{x}_m\right] \tag{3.3.19}$$

in $\mathbb{M}_+([0, \infty]^{m+1} \backslash \{\boldsymbol{0}\})$ for $m \geq 0$, and that Condition 3.3.5 holds with respect to P_H.

(a) If $G(\{0\}) = 0$, and K satisfies Condition 3.3.3, then

$$\sum_{j=0}^{\infty} \epsilon_{\left(\frac{j}{n}, \frac{X_j}{b_n}\right)} \Rightarrow \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \epsilon_{(qt_k,\, i_k \xi_k(j))} \qquad \text{in } \mathbb{M}_p([0, \infty) \times (0, \infty]) \quad \text{as } n \to \infty.$$

(b) If $G(\{0\}) > 0$, and either $y_0(t) \equiv 0$ is an extremal boundary for K, or K satisfies Condition 3.3.4, then

$$\sum_{j=0}^{\infty} \epsilon_{\left(\frac{j}{n}, \frac{X_j}{b_n}\right)} \Rightarrow \sum_{k=0}^{\infty} \sum_{j=0}^{\tau^*_{k+1}-1} \epsilon_{(qt_k,\, i_k \xi_k(j))} \qquad \text{in } \mathbb{M}_p([0, \infty) \times (0, \infty]) \quad \text{as } n \to \infty,$$

where the $\{\tau^*_k\}$ are iid Geometric random variables with parameter $G(\{0\})$.

3.4 Proof of Theorem 3.3.1

Recall that

(i) $\{\boldsymbol{X}_k\}$ are iid copies of a Markov chain $\boldsymbol{X}' \sim K \in D(G)$, with initial distribution H satisfying $tH(b(t)\cdot) \xrightarrow{v} \nu_\alpha$ in $\mathbb{M}_+(0,\infty]$, where $b(t) \to \infty$, and $\{\tilde{\tau}_k(t)\}$ are the respective extremal boundary downcrossing times;

(ii) $\{\boldsymbol{\xi}_k\}$ are iid copies of the multiplicative random walk $\boldsymbol{\xi}$ with associated stopping times $\{\tau_k^*\}$;

(iii) $\zeta = \sum \epsilon_{(t_k,\, i_k)}$ is PRM on $[0,\infty) \times (0,\infty]$ with mean measure $\mathbb{LEB} \times \nu_\alpha$, independent of the $\{\boldsymbol{\xi}_k\}$; and

(iv) $\{S_k\}$ is the renewal process given by (3.3.3) with finite mean interarrival time q.

For convenience, write $\boldsymbol{X}_{k,m}^{(t)} = (X_{k0}^{(t)}, \ldots, X_{km}^{(t)})$ and $\boldsymbol{\xi}_{k,m} = (\xi_k(0), \ldots, \xi_k(m))$.

Proof of Theorem 3.3.1. (a) Recall that, under our assumptions, the convergence (3.2.2) (p. 48) of $\boldsymbol{X}'$ takes place on the space $\mathbb{E} = (0,\infty] \times [0,\infty]^m$, with limit measure μ given by (3.2.1). Applying the Poisson transform [77, Corollary 6.1, p. 183], we obtain

$$\sum_{k=0}^{\infty} \epsilon_{\left(\frac{k}{n},\, \frac{\boldsymbol{X}_{k,m}^{(bn)}}{bn}\right)} = \sum_{k=0}^{\infty} \epsilon_{\left(\frac{k}{n},\, \frac{X_{k0}^{(bn)}}{bn}, \ldots, \frac{X_{km}^{(bn)}}{bn}\right)} \Rightarrow \mathrm{PRM}(\mathbb{LEB} \times \mu) \quad \text{in } \mathbb{M}_p([0,\infty) \times \mathbb{E}).$$

Observing that

$$\begin{aligned}\mathsf{P}\left[i_k(\xi_k(1), \ldots, \xi_k(m)) \in A \,\middle|\, (t_k, i_k),\ \left\{(t_\ell, i_\ell, \boldsymbol{\xi}_{\ell,m})\,;\, \ell \neq k\right\}\right] \\ = \mathsf{P}\left[i_k(\xi_k(1), \ldots, \xi_k(m)) \in A \,\middle|\, (t_k,\, i_k)\right] =: \kappa((t_k,\, i_k)\,,\, A),\end{aligned}$$

it follows by [77, Proposition 5.6, p. 144] that

$$\sum_{k=0}^{\infty} \epsilon_{(t_k,\, i_k \boldsymbol{\xi}_{k,m})} = \sum_{k=0}^{\infty} \epsilon_{(t_k,\, i_k,\, i_k\xi_k(1), \ldots,\, i_k\xi_k(m))} \sim \mathrm{PRM}(\mu') \qquad \text{on } \mathbb{M}_p([0,\infty) \times \mathbb{E}),$$

where

$$\mu'(ds, dx_0, d\boldsymbol{x}_m) = \mathbb{LEB}(ds)\nu_\alpha(dx_0)\kappa((s, x_0), d\boldsymbol{x}_m) = \mathbb{LEB}(ds)\mu(dx_0, d\boldsymbol{x}_m)$$

by (3.2.1). Therefore,

$$\vartheta_n = \sum_{k=0}^{\infty} \epsilon_{\left(\frac{k}{n},\, \frac{\boldsymbol{X}_{k,m}^{(b_n)}}{b_n}\right)} \Rightarrow \sum_{k=0}^{\infty} \epsilon_{(t_k,\, i_k \boldsymbol{\xi}_{k,m})} = \vartheta \qquad \text{in } \mathbb{M}_p([0,\infty) \times \mathbb{E}). \tag{3.4.1}$$

Next, we rescale the time axis so that the points fall at the epochs S_k. Let $N(t) = \sum_k \epsilon_{S_k}[0,t] = \inf\{k : S_k > t\}$ be the associated counting function. Inverting, we obtain

$$N^{\rightarrow}(t) = \inf\{s : N(s) > [t]\} = \inf\{s : S_{[t]} \leq s\} = S_{[t]}.$$

Define $N_n(\cdot) = n^{-1} N^{\rightarrow}(n\,\cdot)$, so that $S_k/n = N_n(k/n)$. Note that N_n is a random element of $D^{\uparrow}[0,\infty)$, the subspace of non-decreasing elements of $D[0,\infty)$. Since

$$N_n(t) = \frac{[nt]}{n} \frac{S_{[nt]}}{[nt]} \longrightarrow t \cdot q\,, \qquad t \geq 0, \quad \text{a.s.}$$

by the Strong Law of Large Numbers, it follows that $N_n \to c$ a.s. in $D^{\uparrow}[0,\infty)$, where $c(t) = qt$. We transform the timepoints via the mapping $T : D^{\uparrow}[0,\infty) \times \mathbb{M}_+([0,\infty) \times \mathbb{E}) \to \mathbb{M}_+([0,\infty) \times \mathbb{E})$ given by $T(x,m) = \tilde{m}$, where $\tilde{m}(f)$ is the measure defined by

$$\tilde{m}(f) = \iint f(x(u), v) m(du, dv), \qquad f \in \mathcal{C}_K^+([0,\infty) \times \mathbb{E}). \tag{3.4.2}$$

Applying [77, Proposition 3.1, p. 57] to (3.4.1), we have $(N_n, \vartheta_n) \Rightarrow (c, \vartheta)$ in $D^{\uparrow}[0,\infty) \times \mathbb{M}_p([0,\infty) \times \mathbb{E})$. Since T is a.s. continuous at (c,ϑ) (Lemma 3.4.1, p. 87), the Continuous Mapping Theorem yields, in $\mathbb{M}_p([0,\infty) \times \mathbb{E})$,

$$\eta_n' = \sum_{k=0}^{\infty} \epsilon_{\left(\frac{S_k}{n},\, \frac{\boldsymbol{X}_{k,m}^{(b_n)}}{b_n}\right)} = T(N_n, \vartheta_n) \Rightarrow T(c, \vartheta) = \sum_{k=0}^{\infty} \epsilon_{(qt_k,\, i_k \boldsymbol{\xi}_{k,m})} = \eta'.$$

We now stack the components of $\boldsymbol{X}_{k,m}^{(b_n)}$ above their respective timepoints. First, it is necessary to compactify the state space. Let $\Lambda_\delta := [\delta,\infty] \times [0,\infty]^m$, and define the restriction functional $\widehat{T} : \mathbb{M}_p([0,\infty) \times \mathbb{E}) \to \mathbb{M}_p([0,\infty) \times \Lambda_\delta)$ by $\widehat{T}m =$

$m(\cdot\cap[0,\infty)\times\Lambda_\delta)$. By [37, Proposition 3.3], $\widehat{T}$ is almost surely continuous at η' provided $\mathsf{P}[\eta'(\partial([0,\infty)\times\Lambda_\delta))=0]=1$. But

$$\eta'(\partial([0,\infty)\times\Lambda_\delta))=\eta'([0,\infty)\times\{\delta\}\times[0,\infty]^m)=\zeta([0,\infty)\times\{\delta\}),$$

and $\mathsf{E}\,\zeta([0,s]\times\{\delta\})=\mathbb{LEB}[0,s]\,\nu_\alpha\{\delta\}=0$ for all $s>0$, so $\zeta([0,\infty)\times\{\delta\})=0$ w.p. 1. Therefore, in $\mathbb{M}_p([0,\infty)\times\Lambda_\delta)$,

$$\eta_n''=\sum_{k=0}^{\infty}\epsilon_{\left(\frac{S_k}{n},\,\frac{\boldsymbol{X}_{k,m}^{(b_n)}}{b_n}\right)}\mathbf{1}_{\left\{\frac{X_{k0}}{b_n}\ge\delta\right\}}=\widehat{T}(\eta_n')\Rightarrow\widehat{T}(\eta')=\sum_{k=0}^{\infty}\epsilon_{(qt_k,\,i_k\boldsymbol{\xi}_{k,m})}\mathbf{1}_{\{i_k\ge\delta\}}=\eta''.$$

Next, define the stacking functional $T:\mathbb{M}_p([0,\infty)\times\Lambda_\delta)\to\mathbb{M}_p([0,\infty)\times[0,\infty])$ by $Tm=\hat{m}$, where

$$\hat{m}(f)=\iint\Big\{\sum_{j=0}^{m}f(u,v_j)\Big\}m(du,d\boldsymbol{v}),\qquad f\in\mathcal{C}_K^+([0,\infty)\times[0,\infty]).$$

Given such f with support $[0,R]\times[0,\infty]$, the function $\varphi(u,\boldsymbol{v}):=\sum_{j=0}^m f(u,v_j)$ belongs to $\mathcal{C}_K^+([0,\infty)\times\Lambda_\delta)$, since it is clearly non-negative, continuous, and vanishes outside of $[0,R]\times\Lambda_\delta$. Therefore, T is continuous: given $m_n\xrightarrow{v}m$ in $\mathbb{M}_p([0,\infty)\times\Lambda_\delta)$, we have $\hat{m}_n(f)=m_n(\varphi)\to m(\varphi)=\hat{m}(f)$. Consequently, in $\mathbb{M}_p([0,\infty)\times[0,\infty])$,

$$\begin{aligned}\hat{\eta}_n=\sum_{k=0}^{\infty}\sum_{j=0}^{m}\epsilon_{\left(\frac{S_k}{n},\,\frac{X_{kj}^{(b_n)}}{b_n}\right)}\mathbf{1}_{\left\{\frac{X_{k0}}{b_n}\ge\delta\right\}}&=T(\eta_n'')\\ &\Rightarrow T(\eta'')=\sum_{k=0}^{\infty}\sum_{j=0}^{m}\epsilon_{(qt_k,\,i_k\xi_k(j))}\mathbf{1}_{\{i_k\ge\delta\}}=\hat{\eta}.\end{aligned}$$

The next step is to remove the zeros from the stacks. It follows easily that

$$\hat{\eta}_n(\cdot\cap[0,\infty)\times(0,\infty])\Rightarrow\hat{\eta}(\cdot\cap[0,\infty)\times(0,\infty])\quad\text{in }\mathbb{M}_p([0,\infty)\times(0,\infty])\tag{3.4.3}$$

by noting that any $f\in\mathcal{C}_K^+([0,\infty)\times(0,\infty])$ extends to $\bar{f}\in\mathcal{C}_K^+([0,\infty)\times[0,\infty])$ by setting $\bar{f}(s,0)=0$ for $s\ge0$. Moreover, recalling $\{\tilde{\tau}_{k+1}(t)\}$ and $\{\tau_{k+1}^*\}$, the first hitting times of 0 by $\{\boldsymbol{X}_k^{(t)}\}$ and $\{\boldsymbol{\xi}_k\}$ respectively, put

$$\sigma_{k+1}(t)=\tilde{\tau}_{k+1}(t)\wedge(m+1)\qquad\text{and}\qquad\sigma_{k+1}^*=\tau_{k+1}^*\wedge(m+1),\qquad k\ge0.$$

The convergence (3.4.3) becomes, in $\mathbb{M}_p([0,\infty) \times (0,\infty])$,

$$\tilde{\eta}_n = \sum_{k=0}^{\infty} \sum_{j=0}^{\sigma_{k+1}(b_n)-1} \epsilon_{\left(\frac{S_k}{n},\, \frac{X_{kj}}{b_n}\right)} \mathbf{1}_{\left\{\frac{X_{k0}}{b_n} \geq \delta\right\}} \Rightarrow \sum_{k=0}^{\infty} \sum_{j=0}^{\sigma^*_{k+1}-1} \epsilon_{(qt_k,\, i_k \xi_k(j))} \mathbf{1}_{\{i_k \geq \delta\}} = \tilde{\eta}. \tag{3.4.4}$$

Now let us spread out the stacks of the $\tilde{\eta}_n$ in time. We claim that

$$\tilde{\eta}^*_n = \sum_{k=0}^{\infty} \sum_{j=0}^{\sigma_{k+1}(b_n)-1} \epsilon_{\left(\frac{S_k+j}{n},\, \frac{X_{kj}}{b_n}\right)} \mathbf{1}_{\left\{\frac{X_{k0}}{b_n} \geq \delta\right\}} \Rightarrow \tilde{\eta} \qquad \text{in } \mathbb{M}_p([0,\infty) \times (0,\infty]). \tag{3.4.5}$$

This will follow from Slutsky's theorem provided $d_v(\tilde{\eta}^*_n, \tilde{\eta}_n) \xrightarrow{P} 0$, where d_v is the vague metric on $\mathbb{M}_p([0,\infty) \times (0,\infty])$, which will hold if $\mathsf{P}[|\tilde{\eta}^*_n(f) - \tilde{\eta}_n(f)| > \gamma] \to 0$ for any $f \in \mathcal{C}_K^+([0,\infty) \times (0,\infty])$ and $\gamma > 0$. For such f with support $[0,R] \times [M,\infty]$, we have

$$\begin{aligned}
&\mathsf{P}\left[|\tilde{\eta}^*_n(f) - \tilde{\eta}_n(f)| > \gamma\right] \\
&\quad = \mathsf{P}\Bigg[\Bigg| \sum_{k=0}^{\infty} \sum_{j=0}^{\sigma_{k+1}(b_n)-1} f\left(\frac{S_k+j}{n}, \frac{X_{kj}}{b_n}\right) \mathbf{1}_{\left\{\frac{X_{k0}}{b_n} \geq \delta\right\}} \\
&\qquad\qquad - \sum_{k=0}^{\infty} \sum_{j=0}^{\sigma_{k+1}(b_n)-1} f\left(\frac{S_k}{n}, \frac{X_{kj}}{b_n}\right) \mathbf{1}_{\left\{\frac{X_{k0}}{b_n} \geq \delta\right\}} \Bigg| > \gamma \Bigg] \\
&\quad \leq \mathsf{P}\Bigg[\sum_{k=0}^{\infty} \sum_{j=1}^{\sigma_{k+1}(b_n)-1} \left| f\left(\frac{S_k+j}{n}, \frac{X_{kj}}{b_n}\right) - f\left(\frac{S_k}{n}, \frac{X_{kj}}{b_n}\right) \right| \\
&\qquad\qquad \mathbf{1}_{\{[0,R]\times[M,\infty]\}}\left(\frac{S_k}{n}, \frac{X_{kj}}{b_n}\right) \mathbf{1}_{\left\{\frac{X_{k0}}{b_n} \geq \delta\right\}} > \gamma \Bigg].
\end{aligned}$$

Now, given $\rho > 0$, let $v > 0$ be such that $|f(x) - f(y)| < \rho$ whenever $\|x - y\| < v$. Then for n large enough that $m/n < v$, we have

$$\begin{aligned}
&\sum_{k=0}^{\infty} \sum_{j=1}^{\sigma_{k+1}(b_n)-1} \left| f\left(\frac{S_k+j}{n}, \frac{X_{kj}}{b_n}\right) - f\left(\frac{S_k}{n}, \frac{X_{kj}}{b_n}\right) \right| \mathbf{1}_{\{[0,R]\times[M,\infty]\}}\left(\frac{S_k}{n}, \frac{X_{kj}}{b_n}\right) \mathbf{1}_{\left\{\frac{X_{k0}}{b_n} \geq \delta\right\}} \\
&\qquad < \rho \cdot \tilde{\eta}_n([0,R] \times [M,\infty]),
\end{aligned}$$

implying that

$$\limsup_{n\to\infty} \mathsf{P}\left[|\tilde{\eta}_n^*(f) - \tilde{\eta}_n(f)| > \gamma\right] \le \limsup_{n\to\infty} \mathsf{P}\left[\tilde{\eta}_n([0,R]\times[M,\infty]) > \gamma\rho^{-1}\right]$$
$$\le \mathsf{P}\left[\tilde{\eta}([0,R]\times[M,\infty]) \ge \gamma\rho^{-1}\right]$$

by (3.4.4), provided

$$\mathsf{P}\left[\tilde{\eta}(\partial([0,R]\times[M,\infty])) = 0\right] = 1. \tag{3.4.6}$$

But this probability is bounded by

$$\rho\gamma^{-1}\cdot \mathsf{E}\,\tilde{\eta}\big([0,R]\times[M,\infty]\big) \le \rho\gamma^{-1}m\cdot \mathsf{E}\,\zeta\big([0,q^{-1}R]\times[\delta,\infty]\big) = \rho\gamma^{-1}mq^{-1}R\delta^{-\alpha},$$

so (3.4.5) follows by letting $\rho\to 0$.

It remains to verify (3.4.6). On the one hand,

$$\mathsf{P}[\tilde{\eta}(\{R\}\times[M,\infty]) > 0] \le \mathsf{P}[\zeta(\{R/q\}\times[\delta,\infty]) > 0] = 0,$$

since $\mathsf{E}\,\zeta(\{R/q\}\times[\delta,\infty]) = \delta^{-\alpha}\mathbb{LEB}\{R/q\} = 0$. On the other,

$$\begin{aligned}\tilde{\eta}\big([0,R]\times\{M\}\big) &= \sum_{k=0}^{\infty}\epsilon_{(t_k,i_k)}[0,R/q]\times[\delta,\infty]\sum_{j=0}^{m}\mathbf{1}_{\{\xi_k(j)>0\}}\,\mathbf{1}_{\{i_k=M/\xi_k(j)\}}\\ &= \sum_{j=0}^{m}\sum_{k=0}^{\infty}\epsilon_{(t_k,i_k)}[0,R/q]\times\{M/\xi_k(j)\}\,\mathbf{1}_{\{0<\xi_k(j)\le M/\delta\}}\\ &\le \sum_{j=0}^{m}\sum_{k=0}^{\infty}\epsilon_{(t_k,i_k)}[0,R/q]\times\Bigg\{\bigcup_{\ell:\,0<\xi_\ell(j)\le M/\delta}\{M/\xi_\ell(j)\}\Bigg\}.\end{aligned}$$

The expectation of the j-th summand is

$$\begin{aligned}&\mathsf{E}\,\zeta\Bigg([0,R/q]\times\Bigg\{\bigcup_{\ell:\,0<\xi_\ell(j)\le M/\delta}\{M/\xi_\ell(j)\}\Bigg\}\Bigg)\\ &\qquad = \mathsf{E}\,\mathsf{E}\Bigg[\zeta\Bigg([0,R/q]\times\Bigg\{\bigcup_{\ell:\,0<\xi_\ell(j)\le M/\delta}\{M/\xi_\ell(j)\}\Bigg\}\Bigg)\,\Bigg|\,\big\{\xi_\ell(j)\,;\,\ell\ge 0\big\}\Bigg] = 0\end{aligned}$$

since

$$\mathsf{E}\,\zeta\bigg([0,R/q]\times\bigg\{\bigcup_{\ell:\,0<u_\ell\le M/\delta}\{M/u_\ell\}\bigg\}\bigg)=\sum_{\ell:\,0<u_\ell\le M/\delta}\mathsf{E}\,\zeta\big([0,R/q]\times\{M/u_\ell\}\big)$$
$$=\sum_{\ell:\,0<u_\ell\le M/\delta}(R/q)\nu_\alpha\{M/u_\ell\}=0,$$

and the $\{\xi_\ell(j)\}$ are independent of ζ. Hence, $\mathsf{E}\,\tilde{\eta}([0,R]\times\{M\})=0$, leading to (3.4.6).

Finally, we remove the restriction in m on the stacks. This is accomplished by showing that

$$\lim_{m\to\infty}\mathsf{P}\left[d_v(\tilde{\eta},\eta)>\gamma\right]=0\quad\text{and}\quad\lim_{m\to\infty}\limsup_{n\to\infty}\mathsf{P}\left[d_v(\tilde{\eta}^*_n,\eta_n)>\gamma\right]=0 \tag{3.4.7}$$

for any $\gamma>0$ ([77] Theorem 3.4, p. 56). Let $f\in\mathcal{C}_K^+([0,\infty)\times(0,\infty])$ with support $[0,R]\times[M,\infty]$. Taking $\delta<a<\infty$, we can write

$$|\tilde{\eta}(f)-\eta(f)|=\sum_{k=0}^{\infty}\sum_{j=m+1}^{\infty}f(qt_k,i_k\xi_k(j))\cdot\big(\mathbf{1}_{\{\delta\le i_k<a\}}+\mathbf{1}_{\{i_k\ge a\}}\big).$$

Hence,

$$\mathsf{P}\left[|\tilde{\eta}(f)-\eta(f)|>\gamma\right]\le\mathsf{P}\left[\sum_{k=0}^{\infty}\sum_{j=m+1}^{\infty}f(qt_k,i_k\xi_k(j))\,\mathbf{1}_{\{\delta\le i_k<a\}}>\gamma/2\right]$$
$$+\mathsf{P}\left[\sum_{k=0}^{\infty}\sum_{j=m+1}^{\infty}f(qt_k,i_k\xi_k(j))\,\mathbf{1}_{\{i_k\ge a\}}>\gamma/2\right].$$

Writing $\xi_k^*(m)=\sup_{j\ge m+1}\xi_k(j)$ for $k\ge0$, the first term is bounded by

$$\mathsf{P}\left[\sum_{k=0}^{\infty}\epsilon_{(t_k,i_k)}[0,R/q]\times[\delta,a)\sum_{j=m+1}^{\infty}\mathbf{1}_{\{\xi_k(j)>Ma^{-1}\}}>0\right]$$
$$\le\mathsf{P}\left[\zeta_m'\big([0,R/q]\times[\delta,\infty]\times(Ma^{-1},\infty)\big)>0\right],$$

where

$$\zeta_m'=\sum_{k=0}^{\infty}\epsilon_{(t_k,i_k,\xi_k^*(m))}\sim\mathrm{PRM}\big(\mathbb{LEB}\times\nu_\alpha\times\mathsf{P}\,[\sup_{j\ge m+1}\xi(j)\in\cdot\,]\big)$$

in $\mathbb{M}_p([0,\infty)\times(0,\infty]\times[0,\infty])$, recalling that the $\{\boldsymbol{\xi}_k\}$ are iid and independent of ζ. Therefore, $\mathsf{P}[\zeta'_m([0,R/q]\times[\delta,\infty]\times(Ma^{-1},\infty))>0]=1-\exp\{-\lambda\}$, where

$$\begin{aligned}\lambda &= \mathbb{LEB}[0,R/q]\cdot\nu_\alpha[\delta,\infty]\cdot\mathsf{P}\big[\sup_{j\geq m+1}\xi(j)>Ma^{-1}\big]\\ &= Rq^{-1}\delta^{-\alpha}\,\mathsf{P}\big[\sup_{j\geq m+1}\xi(j)>Ma^{-1}\big]\longrightarrow 0\end{aligned}$$

as $m\to\infty$ by (3.2.12), a consequence of Condition 3.2.1. The second is at most

$$\begin{aligned}\mathsf{P}\big[\zeta([0,R/q]\times[a,\infty])>0\big] &= 1-\exp\big\{-\mathsf{E}\,\zeta([0,R/q]\times[a,\infty])\big\}\\ &= 1-\exp\{-Rq^{-1}a^{-\alpha}\}.\end{aligned}$$

Letting $a\to\infty$ establishes the first limit in (3.4.7). For the second limit, observe that

$$\begin{aligned}\mathsf{P}\big[|\tilde\eta_n^*(f)-\eta_n(f)|>\gamma\big] &= \mathsf{P}\Bigg[\sum_{k=0}^{\infty}\sum_{j=m+1}^{\tilde\tau_{k+1}(b_n)-1} f\Big(\frac{S_k+j}{n},\frac{X_{kj}}{b_n}\Big)\mathbf{1}_{\left\{\frac{X_{k0}}{b_n}\geq\delta\right\}}>\gamma\Bigg]\\ &\leq \mathsf{P}\Bigg[\sum_{k=0}^{2Rn-1}\sum_{j=m+1}^{\infty} f\Big(\frac{S_k+j}{n},\frac{X_{kj}^{(b_n)}}{b_n}\Big)\mathbf{1}_{\left\{\frac{X_{k0}}{b_n}\geq\delta\right\}}>0\Bigg]\\ &\qquad+\mathsf{P}\Bigg[\sum_{k=2Rn}^{\infty}\sum_{j=m+1}^{\infty} f\Big(\frac{S_k+j}{n},\frac{X_{kj}^{(b_n)}}{b_n}\Big)\mathbf{1}_{\left\{\frac{X_{k0}}{b_n}\geq\delta\right\}}>0\Bigg].\end{aligned}$$

The first term is bounded by

$$\begin{aligned}&\mathsf{P}\Bigg[\bigcup_{k=0}^{2Rn-1}\Bigg(\Big\{\frac{X_{k0}}{b_n}\geq\delta\Big\}\cap\bigcup_{j=m+1}^{\infty}\Big\{\frac{X_{kj}^{(b_n)}}{b_n}\geq M\Big\}\Bigg)\Bigg]\\ &\qquad\leq 2Rn\,\mathsf{P}\Bigg[\frac{X'_0}{b_n}\geq\delta,\ \sup_{j\geq m+1}\frac{X_j'^{(b_n)}}{b_n}\geq M\Bigg],\end{aligned}$$

and

$$\lim_{m\to\infty}\limsup_{n\to\infty} n\,\mathsf{P}\Bigg[\frac{X'_0}{b_n}\geq\delta,\ \sup_{j\geq m+1}\frac{X_j'^{(b_n)}}{b_n}\geq M\Bigg]=0$$

by Condition 3.2.1. The second term is at most

$$\mathsf{P}\Bigg[\frac{S_{2Rn}}{n}\leq R\Bigg]=\mathsf{P}\Bigg[\frac{S_{2Rn}}{2Rn}\leq 1/2\Bigg]\longrightarrow 0$$

as $n \to \infty$, since $S_n/n \to q$ a.s., and $q \geq 1$ by (3.3.5). This establishes (3.4.7), completing the proof.

(b) This amounts to removing the restrictions in δ, under the additional assumptions of joint regular variation (3.2.9) and Condition 3.2.2. We accomplish this by showing that

$$\lim_{\delta \to 0} \mathsf{P}\left[d_v(\eta, \eta^*) > \gamma\right] = 0 \quad \text{and} \quad \lim_{\delta \to 0} \limsup_{n \to \infty} \mathsf{P}\left[d_v(\eta_n, \eta_n^*) > \gamma\right] = 0 \tag{3.4.8}$$

for any $\gamma > 0$. Let $f \in \mathcal{C}_K^+([0,\infty) \times (0,\infty])$ with support $[0,R] \times [M,\infty]$, and note that

$$|\eta(f) - \eta^*(f)| = \sum_{k=0}^{\infty} \sum_{j=0}^{\tau_{k+1}^*-1} f(qt_k,\, i_k \xi_k(j))\, \mathbf{1}_{\{i_k < \delta\}}\,.$$

Hence, writing $\xi_k^* = \sup_{j \geq 1} \xi_k(j)$, we have

$$\begin{aligned}\mathsf{P}\left[|\eta(f) - \eta^*(f)| > \gamma\right] &\leq \mathsf{P}\left[\sum_{k=0}^{\infty} \epsilon_{(t_k,\, i_k)}[0, R/q] \times (0,\delta) \sum_{j=0}^{\infty} \mathbf{1}_{\{i_k \xi_k(j) \geq M\}} > 0\right] \\ &\leq \mathsf{P}\left[\zeta'([0,R/q] \times (0,\delta) \times [M,\infty]) > 0\right]\end{aligned}$$

where $\zeta' = \sum_{k=0}^{\infty} \epsilon_{(t_k,\, i_k,\, i_k \xi_k^*)}$. Since the $\{\boldsymbol{\xi}_k\}$ are iid and independent of ζ, we find that $\zeta' \sim \mathrm{PRM}(\mu')$ on $\mathbb{M}_p[0,\infty) \times (0,\infty] \times [0,\infty]$ with

$$\mu'(ds, dx, dy) = \mathbb{LEB}(dx) \cdot \nu_\alpha(dx) \cdot \mathsf{P}\left[\sup_{j \geq 1} \xi(j) \in x^{-1} dy\right]$$

by [77, Proposition 5.6, p. 144]. Therefore, $\mathsf{P}\left[\zeta'([0,R/q] \times (0,\delta) \times [M,\infty]) > 0\right] = 1 - \exp\{-\lambda\}$, where

$$\begin{aligned}\lambda &= Rq^{-1} \int_{(0,\delta)} \nu_\alpha(dx)\, \mathsf{P}\left[\sup_{j \geq 1} \xi(j) \geq M x^{-1}\right] \\ &\leq Rq^{-1} M^{-\alpha}\, \mathsf{E}\left[\sup_{j \geq 1} \xi(j)^\alpha \cdot \mathbf{1}_{\{\sup_{j \geq 1} \xi(j) > M\delta^{-1}\}}\right]\end{aligned}$$

by Lemma 3.2.1. Applying dominated convergence under (3.2.14) shows $\lambda \to 0$ as

$\delta \downarrow 0$, establishing the first limit in (3.4.8). For the second limit, we have

$$\mathsf{P}\left[|\eta_n(f) - \eta_n^*(f)| > \gamma\right] = \mathsf{P}\left[\sum_{k=0}^{\infty} \sum_{j=0}^{\tilde{\tau}_{k+1}(b_n)-1} f\left(\frac{S_k + j}{n}, \frac{X_{kj}}{b_n}\right) \mathbf{1}_{\left\{\frac{X_{k0}^{(b_n)}}{b_n} < \delta\right\}} > \gamma\right]$$

$$\leq \mathsf{P}\left[\sum_{k=0}^{2Rn-1} \sum_{j=0}^{\tilde{\tau}_{k+1}(b_n)-1} f\left(\frac{S_k + j}{n}, \frac{X_{kj}}{b_n}\right) \mathbf{1}_{\left\{\frac{X_{k0}^{(b_n)}}{b_n} < \delta\right\}} > 0\right]$$

$$+ \mathsf{P}\left[\sum_{k=2Rn}^{\infty} \sum_{j=0}^{\tilde{\tau}_{k+1}(b_n)-1} f\left(\frac{S_k + j}{n}, \frac{X_{kj}}{b_n}\right) \mathbf{1}_{\left\{\frac{X_{k0}^{(b_n)}}{b_n} < \delta\right\}} > 0\right].$$

As above, the second term is at most $\mathsf{P}[S_{2Rn}/n \leq R] \to 0$ as $n \to \infty$. The first term is bounded by

$$\mathsf{P}\left[\bigcup_{k=0}^{2Rn-1} \left(\left\{\frac{X_{k0}^{(b_n)}}{b_n} < \delta\right\} \cap \bigcup_{j=1}^{\tilde{\tau}_{k+1}(b_n)-1} \left\{\frac{X_{kj}}{b_n} \geq M\right\}\right)\right]$$

$$\leq 2Rn\,\mathsf{P}\left[\frac{X_0^{(b_n)}}{b_n} < \delta, \sup_{j \geq 1} \frac{X_j^{(b_n)}}{b_n} \geq M\right]$$

$$\leq 2Rn\,\mathsf{P}\left[\frac{X_0^{(b_n)}}{b_n} < \delta, \sup_{1 \leq j \leq m_0} \frac{X_j^{(b_n)}}{b_n} \geq M\right]$$

$$+ 2Rn\,\mathsf{P}\left[\frac{X_0^{(b_n)}}{b_n} < \delta, \sup_{j \geq m_0+1} \frac{X_j^{(b_n)}}{b_n} \geq M\right]$$

$$= A_n(\delta) + B_n(\delta),$$

with m_0 as in Condition 3.2.2. For the first term, we have

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} A_n(\delta) = \lim_{\delta \downarrow 0} 2R\mu^*([0, \delta] \times ([0, M)^{m_0})^{\mathrm{c}}) = 0$$

by (3.2.9). For the second, $\lim_{\delta \downarrow 0} \limsup_{n \to \infty} B_n(\delta) = 0$ by Condition 3.2.2. This establishes (3.4.8). □

The following lemma verifies the continuity of the map T defined in (3.4.2).

Lemma 3.4.1. *The mapping* $T : D^{\uparrow}[0, \infty) \times \mathbb{M}_+([0, \infty) \times \mathbb{E}) \to \mathbb{M}_+([0, \infty) \times \mathbb{E})$ *given by* $T(x, m) = \tilde{m}$*, where* $\tilde{m}(f)$ *is the measure defined by*

$$\tilde{m}(f) = \iint f(x(u), v)m(du, dv), \qquad f \in \mathcal{C}_K^+([0, \infty) \times \mathbb{E}),$$

is continuous at (x, m) whenever the function $x = x(t)$ is continuous.

Proof. (a) Suppose $x_n \to x_0$ in $D^{\uparrow}[0, \infty)$ (with respect to the Skorohod topology), where x_0 is continuous, and $m_n \stackrel{v}{\to} m_0$ in $\mathbb{M}_+([0, \infty) \times \mathbb{E})$. Let $f \in \mathcal{C}_K^+([0, \infty) \times \mathbb{E})$ with support contained in $[0, R] \times B$. We show that $\tilde{m}_n(f) \to \tilde{m}_0(f)$. Write $f_n(u, v) = f(x_n(u), v)$, $n \geq 0$. The f_n are supported on $x_n^{-1}([0, R]) \times B$, and $x_n^{-1}([0, R]) = [0, x_n^{\to}(R)]$, where $x_n^{\to}$ is the right-continuous inverse of x_n. We now argue that the f_n, $n \geq 0$, have a common compact support. Indeed, we have $x_n^{\to} \to x_0^{\to}$ pointwise, so $x_n^{\to}(R) \to x_0^{\to}(R)$. Thus, for large n, $[0, x_n^{\to}(R)] \times B \subset [0, x_0^{\to}(R) + 1] \times B$; without loss of generality, $m_0(\partial([0, x_0^{\to}(R) + 1] \times B)) = 0$. Furthermore, $f_n \to f_0$ uniformly: suppose $(u_n, v_n) \to (u_0, v_0) \in [0, \infty) \times \mathbb{E}$. Then $x_n(u_n) \to x_0(u_0)$ since x_0 is continuous, and so $f(x_n(u_n), v_n) \to f(x_0(u_0), v_0)$ by the continuity of f. Consequently, $\tilde{m}_n(f) \to \tilde{m}_0(f)$ by Lemma A.0.2 (p. 132). $\square$

3.5 Conclusions and Future Directions

Studying the extremes of a stochastic process via point processes generally requires splitting its sample path into approximately iid blocks in order to draw on results for iid observations. For general stationary processes, this is often accomplished by assuming a mixing condition to control long-range dependence, following the approach of Leadbetter et al. [55]. It is then necessary to choose block sizes that increase slowly enough to maintain approximate independence.

Although such conditions are satisfied by many commonly-studied time series models, often a more specific dependence structure is present. For example, though it is well-known that regenerative processes are strongly mixing, it is much more natural to work with the iid regenerative cycles rather than with blocks of a deterministic size. We have taken this view in following the direction proposed by Rootzén [81].

Rootzén showed that deriving a limit for the process counting the number of exceedances is equivalent to a regular variation-type property for the distribution of the cycle maximum. On the other hand, if the asymptotic behaviour of the finite-dimensional distributions is known, a limit for the process recording the heights of the exceedances may be obtained (see, for example, Davis and Hsing [29]).

We have combined these two components in the context of a Markov chain. If the transition kernel is in a domain of attraction, then we require only additional information about the marginal distributions to obtain a joint regular variation property for the finite-dimensional distributions. Adding to this the (non-asymptotic) regenerative structure leads to the convergence of the "complete" point process recording times and heights of exceedances, where the latter are approximated by the tail chain. Of course, by the nature of the tail chain model discussed in Chapter 2, this approximation is only able to capture exceedances occurring in the extremal components of the cycles. Furthermore, we have made the connection with the asymptotics of the cycle maximum, leading to an expression for the extremal index.

Our approach is applicable in more general circumstances, where it is reasonable to assume a regenerative structure. Basrak and Segers [8] develop the idea of a "tail process" describing the limit measure of regularly varying finite-dimensional distributions for general stationary processes, using it to determine the heights of the points in a limiting point process similar to ours. This suggests that detailed information regarding exceedance heights is available in much broader generality than we have assumed in this chapter, and that the techniques that we have used could be applied to general regularly varying regenerative processes. Furthermore, our results show that stationarity is not a crucial feature in the development. The main ingredients are information about long-range dependence and asymptotics of the tails of joint distributions. Stationarity provides a convenient framework in which to control these, but the overall equality of distributions is not strictly necessary. Moreover, for the sake of clarity, we have worked with Markov chains that are positive recurrent. In general, Harris recurrent chains that do not possess an explicit atom display a 1-dependent regenerative structure, in which adjacent cycles are not necessarily independent. A natural extension of our results would be to consider point process limits in this case.

CHAPTER 4

CONDITIONAL EXTREME VALUE MODELING

4.1 Overview

The classical approach to extreme value modelling for multivariate data is to assume that the joint distribution belongs to a multivariate domain of attraction. In particular, this requires that each marginal distribution be individually attracted to a univariate extreme value distribution. The domain of attraction condition may be phrased conveniently in terms of regular variation of the joint distribution on an appropriate cone; see Das and Resnick [25, Proposition 4.1].

A more flexible model for data realizations of a random vector was proposed by Heffernan and Tawn [45], under which not all the components are required to belong to an extremal domain of attraction. Such a model accomodates varying degrees of asymptotic dependence between pairs of components. Instead of starting from the joint distribution, Heffernan and Tawn assumed the existence of an asymptotic approximation to the conditional distribution of the random vector given one of the components, as that component becomes extreme. Combined with the knowledge that the conditioning component belongs to a univariate domain of attraction, this leads to an approximation for the joint distribution, given that one component is extreme (e.g., exceeds some high threshold). However, the focus on conditional distributions presents some technical difficulties regarding the choice of version.

This approach was subsequently formalized as the Conditional Extreme Value Model (CEVM) by Heffernan and Resnick [44] and Das and Resnick [25, 26] in terms of regular variation of the joint distributions, but taking place on a smaller cone than the one employed in multivariate extreme value theory. This is related to the concept of hidden regular variation; see Resnick [78].

We return to the formulation of Heffernan and Tawn [45] in terms of conditional distributions, placing it in a more formal context by drawing upon the theory for transition kernels in a domain of attraction developed in Chapter 2. In particular, we assume that the dependence between a pair of random variables (X, Y) is specified by a transition kernel K; this is appropriate, for example, in cases where one variable can be modeled as an explicit function of the other. In order to better fit in with the study of extremes of a random vector, we extend the kernel domain of attraction condition (2.2.5) to accomodate general linear normalization in both the initial state and the distribution of the next state. We examine conditions under which this extends to a CEVM, when combined with a marginal domain of attraction assumption, and we derive explicit formulas for the CEV limit measure in different cases. Also, through a number of revealing examples, we explore the properties of the normalization functions, and technicalities surrounding the choice of version of the conditional distribution and the limit distribution G appearing in (2.2.5).

4.2 Background

We begin by presenting some necessary background material. First, we review the basics of extended regular variation, which features prominently in the formulation of the CEVM, as well as some concepts of univariate extreme value theory. We then introduce the Conditional Extreme Value Model formally and discuss its basic properties.

4.2.1 Technical Preliminaries: Extended Regular Variation

Regular variation plays an important role in the mathematical description of extreme value theory. It is discussed in Section 1.2.2 (p. 6).

Treatment of the CEVM also require the concept of extended regular variation. See de Haan and Ferreira [33, Appendix B.2] for more information. We will say that the pair of functions $a : (0,\infty) \to (0,\infty)$ and $f : (0,\infty) \to \mathbb{R}$ are *extended regularly varying* (ERV) with parameters $\rho, k \in \mathbb{R}$ if

$$\frac{a(tx)}{a(t)} \longrightarrow x^{\rho} \qquad \text{and} \qquad \frac{f(tx) - f(t)}{a(t)} \longrightarrow \psi(x), \qquad x > 0, \tag{4.2.1}$$

as $t \to \infty$, where

$$\psi(x) = \begin{cases} k\rho^{-1}(x^{\rho} - 1) & \rho \neq 0 \\ k \log x & \rho = 0 \end{cases}. \tag{4.2.2}$$

We will write this as $a, f \in \mathrm{ERV}_{\rho,k}$. Thus, $a \in \mathrm{RV}_{\rho}$. A useful identity is

$$\psi(x^{-1}) = -x^{-\rho}\psi(x). \tag{4.2.3}$$

Note that this differs slightly from the usual definition of extended regular variation, which assumes $k = 1$. If $\phi(x) := \lim_{t\to\infty}(f(tx) - f(t))/a(t)$ exists for $x > 0$, then a is necessarily regularly varying, and $\phi \equiv \psi$, the function given in (4.2.2). Also, the convergences in (4.2.1) are locally uniform, implying that

$$\frac{a(tx_t)}{a(t)} \longrightarrow x^{\rho} \qquad \text{and} \qquad \frac{f(tx_t) - f(t)}{a(t)} \longrightarrow \psi(x) \qquad \text{whenever } x_t \to x > 0.$$

Furthermore, if $k \neq 0$ we obtain the following properties depending on the value of ρ. Recall the sign function $\operatorname{sgn}(u) = u/|u|\,\mathbf{1}_{\{u\neq 0\}}$.

- If $\rho > 0$, then $f \cdot \operatorname{sgn}(k) \in \mathrm{RV}_\rho$, and $f(t)/a(t) \to k/\rho$.
- If $\rho < 0$, then $f(\infty) = \lim_{t\to\infty} f(t)$ exists finite, $(f(\infty) - f)\cdot\operatorname{sgn}(k) \in \mathrm{RV}_{-|\rho|}$, and $(f(\infty) - f(t))/a(t) \to k/|\rho|$.
- If $\rho = 0$, i.e., a is slowly varying, then $f \in \Pi(a)$ (see [33, Appendix B.2]). Suppose $k > 0$. Then $f(\infty) \leq \infty$ exists. If $f(\infty) = \infty$, then $f \in \mathrm{RV}_0$ and $f(t)/a(t) \to \infty$. If $f(\infty) < \infty$, then $f(\infty) - f \in \mathrm{RV}_0$, and $(f(\infty) - f(t))/a(t) \to \infty$. If $k < 0$, then $-f$ has these properties.

4.2.2 Technical Preliminaries: Domains of Attraction

For $\gamma \in \mathbb{R}$, define $\mathbb{E}_\gamma = \{x \in \mathbb{R} : 1 + \gamma x > 0\}$. Observe that

$$\mathbb{E}_\gamma = \begin{cases} (-\gamma^{-1}, \infty) & \gamma > 0 \\ (-\infty, \infty) & \gamma = 0 \\ (-\infty, |\gamma|^{-1}) & \gamma < 0 \end{cases} \cdot \tag{4.2.4}$$

The distribution F_Y of a random variable Y is in the *domain of attraction* of an extreme value distribution G_γ for some $\gamma \in \mathbb{R}$, written $F \in D(G_\gamma)$, if there exist functions $a(t) > 0$ and $b(t) \in \mathbb{R}$ such that

$$F^t\big(a(t)y + b(t)\big) \longrightarrow G_\gamma([-\infty, y])$$

weakly as $t \to \infty$, where $G_\gamma([-\infty, y]) = \exp\{-(1+\gamma y)^{-1/\gamma}\}$ for $y \in \mathbb{E}_\gamma$. See de Haan and Ferreira [33] or Resnick [76] for details. This can be reformulated in terms of the tail of the distribution F as

$$t\,\mathsf{P}\left[\frac{Y - b(t)}{a(t)} > y\right] \longrightarrow (1+\gamma y)^{-1/\gamma}, \qquad y \in \mathbb{E}_\gamma. \tag{4.2.5}$$

If $\gamma = 0$, we interpret the limit as e^{-y}.

It is well-known (for example, see [33, Theorem 1.1.6, p. 10]) that if (4.2.5) holds for some functions a and b, then it holds for

$$b(t) = \left(\frac{1}{1 - F_Y}\right)^{\leftarrow}(t) = F_Y^{\leftarrow}\big(1 - t^{-1}\big), \tag{4.2.6}$$

where $g^{\leftarrow}$ denotes the left-continuous inverse of a nondecreasing function g. Hence, by inversion, (4.2.5) implies that

$$\frac{b(tx) - b(t)}{a(t)} \to \frac{x^{\gamma} - 1}{\gamma}\,\mathbf{1}_{\{\gamma \neq 0\}} + \log x\,\mathbf{1}_{\{\gamma = 0\}}, \tag{4.2.7}$$

i.e., $a, b \in \mathrm{ERV}_{\gamma,1}$. Furthermore, if functions $\tilde{a} > 0$ and $\tilde{b} \in \mathbb{R}$ on $(0, \infty)$ are *asymptotically equivalent* to a, b, i.e., they satisfy

$$\frac{\tilde{a}(t)}{a(t)} \longrightarrow 1 \qquad \text{and} \qquad \frac{\tilde{b}(t) - b(t)}{a(t)} \longrightarrow 0 \qquad \text{as } t \to \infty,$$

then (4.2.5) and (4.2.7) hold with a, b replaced by $\tilde{a}, \tilde{b}$. It follows that (4.2.5) is equivalent to $t\,\mathsf{P}[b^{\leftarrow}(Y) > ty] \to y^{-1}$ for $y > 0$, i.e., $1 - F_{b^{\leftarrow}(Y)} \in \mathrm{RV}_{-1}$. This is known as standardization (see [76, Chapter 5]). We will say that Y is in the *standardized domain of attraction*, and write $F \in D(G_1^*)$, if

$$t\,\mathsf{P}[Y > ty] \longrightarrow y^{-1}, \qquad y > 0.$$

4.2.3 The Conditional Extreme Value Model

Denote by $\overline{\mathbb{E}}_{\gamma}$ the closure on the right of the interval $\mathbb{E}_{\gamma}$. A bivariate random vector (X, Y) on $\mathbb{R}^2$ follows a *Conditional Extreme Value Model* (CEVM) if there exist a non-null Radon measure μ on $[-\infty, \infty] \times \overline{\mathbb{E}}_{\gamma}$, and functions $a(t), \alpha(t) > 0$, $b(t), \beta(t) \in \mathbb{R}$, such that, as $t \to \infty$,

$$t\,\mathsf{P}\left[\left(\frac{X - \beta(t)}{\alpha(t)}, \frac{Y - b(t)}{a(t)}\right) \in \cdot\,\right] \xrightarrow{v} \mu(\cdot) \qquad \text{in } \mathbb{M}_+([-\infty, \infty] \times \overline{\mathbb{E}}_{\gamma}), \tag{4.2.8}$$

where μ satisfies the *conditional non-degeneracy* conditions: for each $y \in \mathbb{E}_\gamma$,

$$\begin{aligned} &\mu([-\infty, x] \times (y, \infty]) \text{ is not a degenerate distribution in } x; \\ &\mu([-\infty, x] \times (y, \infty]) < \infty. \end{aligned} \tag{4.2.9}$$

It is convenient to choose the normalization such that

$$H(x) := \mu([-\infty, x] \times (0, \infty]) \text{ is a probability distribution on } [-\infty, \infty]. \tag{4.2.10}$$

See Heffernan and Resnick [44] and Das and Resnick [25] for details.

By applying the joint convergence (4.2.8) to rectangles $[-\infty, \infty] \times (y, \infty]$, we see that the distribution of Y is necessarily attracted to G_γ for some γ. Also, an important property is that the functions α, β are ERV for some $\rho, k \in \mathbb{R}$ [44, Proposition 1]. The limit measure μ in (4.2.8) is a product measure if and only if $(\rho, k) = (0, 0)$ [44, Proposition 2].

The finiteness condition $\mu([-\infty, x] \times (y, \infty]) < \infty$ in (4.2.9) has been included in the conditional non-degeneracy conditions as defined in [25, 44]. However, it is redundant, since (4.2.8) implies that

$$\mu([-\infty, x] \times (y, \infty]) \leq \mu([-\infty, \infty] \times (y, \infty]) = (1 + \gamma y)^{-1/\gamma} < \infty$$

for $y \in \mathbb{E}_\gamma$. On the other hand, the development in [25, 44] employs convergence of types arguments which implicitly require there to be no mass on the lines through $\{\infty\}$. Because $Y \in D(G_\gamma)$, that $\mu([-\infty, x] \times \{\infty\}) = 0$ follows directly from (4.2.8). However, we require also that $\mu(\{\infty\} \times (y, \infty]) = 0$, and this is not necessarily implied by (4.2.8). Indeed, Example 4.3.6 presents a case where (4.2.8) holds for two distinct normalizations (which are not asymptotically equivalent), each yielding a different limit measure which satisfies (4.2.9). One of the limit measures has $\mu(\{\infty\} \times (y, \infty]) > 0$.

We therefore propose to rephrase the conditional non-degeneracy conditions for the CEVM as

$$\begin{aligned}&\mu([-\infty, x] \times (y, \infty]) \text{ is not a degenerate distribution in } x;\\&\mu(\{\infty\} \times (y, \infty]) = 0.\end{aligned} \tag{4.2.11}$$

From the discussion above, it is clear that previous results will remain unchanged after replacing (4.2.9) with (4.2.11). Henceforth, we will say that (X, Y) follows a CEVM if (4.2.8) and (4.2.11) are satisfied.

4.3 Standard Case

Let (X, Y) be a random vector on $\mathbb{R}^2$, with dependence specified by a transition kernel K:

$$\mathsf{P}[X \in \cdot \,\big|\, Y = y] = K(y\,,\,\cdot) \qquad y \in \mathbb{R}.$$

Roughly speaking, if Y is in an extremal domain of attraction, and K belongs to the domain of attraction of a probability distribution G, then (X, Y) follows a CEVM.

It is instructive to consider the *standard case* first. This means that $(X, Y) \in [0, \infty)^2$, and $F_Y \in D(G_1^*)$, which we formulate as

$$tF_Y(t\,\cdot) \xrightarrow{v} \nu_1(\cdot) \quad \text{ in } \mathbb{M}_+(0, \infty] \quad \text{ as } t \to \infty, \tag{4.3.1}$$

together with $K \in D(G)$ as defined in (2.2.5), i.e.,

$$K(t\,,\,t[0, x]) \Rightarrow G([0, x]) \quad \text{ on } [0, \infty]. \tag{4.3.2}$$

Here $\Rightarrow$ denotes weak convergence, i.e., convergence at points of continuity of the limit. Henceforth, we write $\xi \sim G$, so that $G(\cdot) = \mathsf{P}[\xi \in \cdot]$.

4.3.1 Standard CEVM Properties

Under these conditions, (X, Y) follow a CEVM provided $G \neq \epsilon_0$, i.e., unit mass at $\{0\}$.

Theorem 4.3.1. *Suppose that the joint distribution of the random vector* (X, Y) *on* $[0,\infty)^2$ *satisfies* (4.3.1) *and* (4.3.2)*, where* G *is a probability distribution on* $[0,\infty)$*. Then*

$$t\,\mathsf{P}\left[(X,Y) \in t\,\cdot\,\right] \xrightarrow{v} \mu(\cdot) \qquad \text{in } \mathbb{M}_+([0,\infty] \times (0,\infty]), \tag{4.3.3}$$

with limit measure μ *given by*

$$\mu([0,x] \times (y,\infty]) = \int_{(y,\infty]} \nu_1(du)\,\mathsf{P}[\xi \leq xu^{-1}], \qquad x, y > 0. \tag{4.3.4}$$

Furthermore, μ *satisfies the conditional non-degeneracy conditions* (4.2.11) *provided* $G \neq \epsilon_0$*.*

Proof. The convergence (4.3.3) is an application of Proposition 3.2.1 (a) (p. 48), with $\alpha = 1$ and $m = 1$, and with Y playing the role of X_0. From (4.3.5) below, we see that $\mu\left([0,x] \times (y,\infty]\right)$ is continuous in x, and not constant provided $G \neq \epsilon_0$. Also, since $\mu((x,\infty] \times (y,\infty]) = \int_{(y,\infty]} \nu_1(du)\,\mathsf{P}[\xi > xu^{-1}]$, that $\mu(\{\infty\} \times (y,\infty]) = 0$ follows from the fact that $G(\{\infty\}) = 0$. Therefore, μ satisfies (4.2.11). □

We now investigate the properties of the limit measure μ. By changing variables, μ can be expressed alternatively as

$$\begin{aligned} \mu([0,x] \times (y,\infty]) &= \frac{1}{x}\int_0^{x/y} du\,\mathsf{P}[\xi \leq u] \\ &= y^{-1}\,\mathsf{P}[\xi \leq x/y] - x^{-1}\,\mathsf{E}\,\xi\,\mathbf{1}_{\{\xi \leq x/y\}}, \end{aligned} \tag{4.3.5}$$

showing that μ is necessarily continuous in x and y. In fact, if G has a density, then so does μ. Note that the continuity in (4.3.5) holds even if G is degenerate, i.e., $G = \epsilon_c$ for some $c > 0$; see Example 4.3.4 (p. 102). Non-degeneracy of G will only become relevant in the non-standard case. Moreover, μ cannot be a product measure [25, Lemma 3.1]. From (4.3.5) we also observe that the y-axis is assigned mass proportional to $G(\{0\})$, as $\mu(\{0\} \times (y, \infty]) = y^{-1} G(\{0\})$, and the mass over vertical slices of space depends on $\mathsf{E}\,\xi$, since $\mu((x, \infty] \times (0, \infty]) = x^{-1}\,\mathsf{E}\,\xi \leq \infty$. In terms of conditional distributions, (4.3.3) implies

$$\mathsf{P}[X \leq tx \,\big|\, Y > t] \Rightarrow H(x) := \mu([0, x] \times (1, \infty]) = \frac{1}{x} \int_0^x \mathsf{P}[\xi \leq u] du.$$

The convergence (4.3.3) extends to standard regular variation on the larger cone $[0, \infty]^2 \backslash \{\mathbf{0}\}$, i.e., (X, Y) are in a bivariate domain of attraction, if and only if $F_X \in D(G_1^*)$ as well [25, Proposition 4.1]. In this case,

$$t\,\mathsf{P}\left[t^{-1}(X, Y) \in [\mathbf{0}, (x, y)]^{\mathrm{c}}\right] \longrightarrow \frac{1}{x}\left(1 + \int_0^{x/y} \mathsf{P}[\xi \leq u] du\right), \tag{4.3.6}$$

implying that $\mathsf{E}\,\xi \leq 1$, and the x-axis receives mass according to $\mu((x, \infty] \times \{0\}) = x^{-1}(1 - \mathsf{E}\,\xi)$.

If $G = \epsilon_0$, then the convergence (4.3.3) will still take place, with limit measure given by $\mu([0, x] \times (y, \infty]) = y^{-1}$. However, conditional non-degeneracy fails, since all the mass lies on the y-axis, so (X, Y) do not follow a standard CEVM. This is in fact a manifestation of asymptotic independence. Indeed,

$$\mathsf{P}[X > tx \,\big|\, Y > t] \rightarrow 0$$

for any x, so, given that Y is extreme (exceeding the threshold $u(t) = t$), it is very unlikely to observe X to be similarly extreme. If the joint distribution of (X, Y) is regularly varying on the larger cone $[0, \infty]^2 \backslash \{\mathbf{0}\}$, then

$$t\,\mathsf{P}\left[t^{-1}(X, Y) \in [\mathbf{0}, (x, y)]^{\mathrm{c}}\right] \longrightarrow x^{-1} + y^{-1},$$

which means that X and Y are asymptotically independent in the usual sense [44, Section 5]. In this case, (X, Y) do not follow a standard CEVM because of degeneracy, although a CEVM may hold if X is normalized differently, as in Section 4.4.

This suggests viewing the parameter $G(\{0\})$ as quantifying the "degree" of asymptotic dependence from Y to X. For example, given Y, we could write X as a mixture

$$X = WX_0 + (1 - W)X_1, \tag{4.3.7}$$

where X_0 and Y are asymptotically independent, X_1 and Y are asymptotically dependent, and $W \sim \text{Bernoulli}(G(\{0\}))$. This relates the canonical form of the update function representation of K (see Section 2.2.3, p. 15). Asymptotic dependence in the reverse direction, given large X, would then be quantified by $1 - \mathsf{E}\,\xi$ if appropriate. The latter phenomenon is hinted at by Segers [84] in his definition of the "back-and-forth tail chain" to approximate stationary Markov chains .

4.3.2 Examples

We now present some examples illuminating properties of the CEVM based on the conditional assumption (4.3.2).

First we show that we can construct a CEVM given any particular choice of G in the conditional assumption (4.3.2). This is similar to [25, Example 8].

Example 4.3.1. Take G to be any probability distribution on $[0, \infty)$ (excluding $G = \epsilon_0$). Let $Y \sim \text{Pareto}(1)$ on $[1, \infty)$, $\xi \sim G$, independent of Y, and put $X = \xi Y$.

A version of the conditional distribution is given by

$$K(y\,,\,\cdot) = \mathsf{P}[X \in \cdot \,\big|\, Y = y] = \mathsf{P}[\xi Y \in \cdot \,\big|\, Y = y] = G(y^{-1}\cdot).$$

Thus, K is the tail kernel associated with G in the sense of Section 2.2.1, and $K(t\,,\, t[0,x]) = G([0,x])$. Consequently, (X,Y) follows a standard CEVM with limit measure as in (4.3.4). In fact, for $x, y > 0$, we have

$$\begin{aligned} \mathsf{P}[X \le x, Y > y] &= \int_{(y,\infty]} K(u\,,\,[0,x]) P[Y \in du] \\ &= \int_{y\vee 1}^{\infty} \mathsf{P}[\xi \le xu^{-1}]u^{-2}du = \frac{1}{x}\int_0^{x\wedge\frac{x}{y}} \mathsf{P}[\xi \le u]du. \end{aligned}$$

Furthermore, (X,Y) belong to a standard bivariate domain of attraction (4.3.6) iff $F_X \in D(G_1^*)$ as well. The marginal distribution of X is given by

$$F_X([0,x]) = \frac{1}{x}\int_0^x \mathsf{P}[\xi \le u]du = H(x),$$

which has density $f_X(x) = x^{-1}\big\{G([0,x]) - H([0,x])\big\}$ for $x \ge 0$. Since

$$\lim_{t\to\infty} t\,\mathsf{P}[X > tx] = \lim_{t\to\infty} \frac{1}{x}\int_0^{tx} \mathsf{P}[\xi > u]du = x^{-1}\,\mathsf{E}\,\xi \quad (\le \infty),$$

(X,Y) belongs to the standard domain of attraction iff $\mathsf{E}\,\xi = 1$.

Using the recipe outlined in Example 4.3.1, we explore the CEVM in a variety of special cases.

Example 4.3.2. If we choose $\xi \sim \text{Exp}(\lambda)$, we have $X = \lambda^{-1}YE$, where $E \sim \text{Exp}(1)$. The limit measure is given by

$$\mu([0,x]\times(y,\infty]) = \frac{1}{x}\int_0^{x/y}(1-e^{-\lambda u})du = \frac{1}{y} - \frac{1}{\lambda x} + \frac{e^{-\lambda x/y}}{\lambda x}.$$

Thus, the marginal distribution of X is given by $F_X(x) = 1 - (\lambda x)^{-1}(1 - e^{-\lambda x})$ with density $f(x) = \lambda^{-1}x^{-2}(1 - e^{-\lambda x}) - x^{-1}e^{-\lambda x}$, and F_X satisfies (4.3.1) iff $\lambda = 1$.

Next, we consider the case where ξ is heavy-tailed.

Example 4.3.3. For $\alpha > 0$ let $\xi \sim \text{Pareto}(\alpha)$, i.e., $G(x) = 1 - x^{-\alpha}$ for $x \geq 1$. The limit measure assigns no mass to $\{(x, y) : 0 \leq x \leq y\}$, and for $x > y > 0$,

$$\mu([0,x] \times (y,\infty]) = \begin{cases} \dfrac{1}{y} - \left(\dfrac{\alpha}{\alpha-1}\right)\dfrac{1}{x} + \dfrac{y^{\alpha-1}}{x^{\alpha}(\alpha-1)} & \alpha > 1 \\ \dfrac{1}{y} - \dfrac{1}{x} - \dfrac{\log x}{x} + \dfrac{\log y}{x} & \alpha = 1 \\ \dfrac{1}{y} + \left(\dfrac{2-\alpha}{1-\alpha}\right)\dfrac{1}{x} + \dfrac{1}{x^{\alpha}y^{1-\alpha}(1-\alpha)} & \alpha < 1 \end{cases} .$$

In particular, $\mathsf{E}\,\xi = \infty$ when $\alpha \leq 1$, and we see that $\mu((x,\infty] \times (y,\infty]) = y^{-1} - \mu([0,x] \times (y,\infty]) \to \infty$ as $y \downarrow 0$.

It is also possible that G be discrete, although the CEVM limit measure μ remains continuous.

Example 4.3.4. Suppose ξ has discrete distribution $\mathsf{P}[\xi = k] = a_k$, $k = 0, 1, \ldots$. In this case, the limit measure is given by

$$\mu([0,x] \times (y,\infty]) = \frac{1}{x}\int_0^{x/y} \Big(\sum_{k=0}^{[u]} a_k\Big) du = \sum_{k=0}^{[x/y]} a_k(y^{-1} - kx^{-1}),$$

which is continuous in x and y, and $F_X(x) = \sum_{k=0}^{[x]} a_k(1 - kx^{-1})$. In particular, if $\mathsf{P}[\xi = c] = 1$ for some $c > 0$, we obtain

$$\mu([0,x] \times (y,\infty]) = (y^{-1} - cx^{-1})\,\mathbf{1}_{\{x > cy > 0\}} .$$

Hence, the conditional non-degeneracy conditions (4.2.9) are satisfied: the limit measure is not degenerate in x, even though G is degenerate.

The final example illustrates how asymptotic indepedence between X and Y is reflected in the limit distribution G.

Example 4.3.5. Consider $Y \sim \text{Pareto}(1)$, and Z independent of Y such that $\mathsf{P}[Z < \infty] = 1$. Take

$$X = Y \vee Z = Y\,\mathbf{1}_{\{Y \geq Z\}} + Z\,\mathbf{1}_{\{Z > Y\}}\,.$$

Given that Y is extreme, it is very unlikely that Z is as extreme as Y, since they are independent. We have

$$K(y\,,\,[0,x]) = \mathsf{P}[Y \leq x, Z \leq x \,|\, Y = y] = \mathsf{P}[Z \leq x]\,\mathbf{1}_{\{x \geq y\}},$$

and so

$$K(t\,,\,t[0,x]) = \mathsf{P}[Z \leq tx]\,\mathbf{1}_{\{x \geq 1\}} \longrightarrow \epsilon_1([0,x]) = \mathbf{1}_{\{x \geq 1\}} = G([0,x]).$$

In other words, when Y is large, $X \approx Y$. As in the previous example, the limit measure is

$$\mu([0,x] \times (y,\infty]) = (y^{-1} - x^{-1})\,\mathbf{1}_{\{x > y > 0\}}\,.$$

On the other hand, consider $X' = Y \wedge Z = Y\,\mathbf{1}_{\{Y < Z\}} + Z\,\mathbf{1}_{\{Z \leq Y\}}$. When Y is large, we should have $X' \approx Z$, so X' is asymptotically independent of Y. Indeed, in this case,

$$K(y\,,\,(x,\infty]) = \mathsf{P}[Y > x, Z > x \,|\, Y = y] = \mathsf{P}[Z > x]\,\mathbf{1}_{\{y > x\}},$$

from which

$$K(t\,,\,t(x,\infty]) = \mathsf{P}[Z > tx]\,\mathbf{1}_{\{x < 1\}} \longrightarrow 0$$

for $x > 0$. Therefore, $G = \epsilon_0$, and the conditional non-degeneracy conditions do not hold.

4.3.3 Counter-Examples

We now investigate the converse to Theorem 4.3.1. As expected, if (X, Y) follows a non-degenerate CEVM as in (4.3.3), and K is a specific version of the conditional distribution $\mathsf{P}[X \in \cdot \,|\, Y = y]$, it does not necessarily follow that there exists a distribution G such that (4.3.2) holds.

We interpret the claim that (4.3.2) fails in two ways. On the one hand, there may exist a probability distribution G on $[0, \infty]$ satisfying (4.3.2) with $G(\{\infty\}) > 0$. On the other hand, it may be possible to obtain two distinct limit distributions down different subsequences $\{t_n\}$ and $\{t'_n\}$.

First, we consider a case where $G(\{\infty\}) > 0$. This example clarifies the difference between (4.2.9) and (4.2.11), emphasizing the importance of the condition $\mu(\{\infty\} \times (y, \infty]) = 0$.

Example 4.3.6. As usual, take $Y \sim \text{Pareto}(1)$ and suppose that

$$X = WY + (1 - W)Y^2,$$

where $W \sim \text{Bernoulli}(p)$ independent of Y. Then

$$K(y\,,\,\cdot) = \mathsf{P}[X \in \cdot \,|\, Y = y] = p\epsilon_y + (1-p)\epsilon_{y^2},$$

so

$$K(t\,,\,t\,\cdot) = p\epsilon_1 + (1-p)\epsilon_t \Rightarrow p\epsilon_1 + (1-p)\epsilon_\infty = G \qquad \text{on } [0, \infty].$$

Indeed, for $0 \leq x < \infty$,

$$K(t\,,\,t[0,x]) = p\epsilon_1([0,x]) + (1-p)\epsilon_t([0,x]) \longrightarrow p\epsilon_1([0,x])$$

and

$$K(t\,,\,t(x,\infty]) \longrightarrow p\epsilon_1((x,\infty]) + 1 - p,$$

showing that $G(\{\infty\}) = 1 - p$.

On the other hand, for $x, y > 0$,

$$\begin{aligned}\mathsf{P}[X \le x, Y > y] &= p\,\mathsf{P}[Y \le x,\, Y > y] + (1-p)\,\mathsf{P}[Y^2 \le x,\, Y > y] \\ &= p\left[\frac{1}{(y \vee 1)} - \frac{1}{x}\right]\mathbf{1}_{\{x \ge (y\vee 1)\}} + (1-p)\left[\frac{1}{(y\vee 1)} - \frac{1}{\sqrt{x}}\right]\mathbf{1}_{\{x \ge (y\vee 1)^2\}},\end{aligned}$$

so for t sufficiently large,

$$\begin{aligned}t\,\mathsf{P}[X \le tx, Y > ty] &= p\left[\frac{1}{y} - \frac{1}{x}\right]\mathbf{1}_{\{x \ge y\}} + (1-p)\left[\frac{1}{y} - \frac{\sqrt{t}}{\sqrt{x}}\right]\mathbf{1}_{\{\sqrt{x}/\sqrt{t} \ge y\}} \\ &\longrightarrow p\left[\frac{1}{y} - \frac{1}{x}\right]\mathbf{1}_{\{x \ge y\}} = \mu([0,x] \times (y,\infty]). \qquad (4.3.8)\end{aligned}$$

The limit clearly satisfies the non-degeneracy conditions (4.2.9). However, μ assigns positive mass to sets $\{\infty\} \times (y, \infty]$:

$$\mu((x,\infty] \times (y,\infty]) = y^{-1} - \mu([0,x] \times (y,\infty]) = \frac{1}{y}\mathbf{1}_{\{x<y\}} + \left[\frac{1-p}{y} + \frac{p}{x}\right]\mathbf{1}_{\{x \ge y\}},$$

showing that $\mu(\{\infty\} \times (y,\infty]) = (1-p)y^{-1}$. Therefore, μ does not satisfy (4.2.11).

In this case, under a different choice of normalization, it is possible to obtain a proper limit G. Indeed, note that

$$K(t\,,\, t^2\cdot) = p\epsilon_{t^{-1}} + (1-p)\epsilon_1 \Rightarrow p\epsilon_0 + (1-p)\epsilon_1 \sim \text{Bernoulli}(1-p),$$

and hence,

$$\begin{aligned}t\,\mathsf{P}[X \le t^2 x, Y > t\cdot y] =& p(y^{-1} - (tx)^{-1})\,\mathbf{1}_{\{x \ge y/t\}} + (1-p)(y^{-1} - x^{-1/2})\,\mathbf{1}_{\{x \ge y\}} \\ &\longrightarrow py^{-1} + (1-p)(y^{-1} - x^{-1/2})\,\mathbf{1}_{\{x \ge y\}}\,.\end{aligned}$$

This limit does satisfy (4.2.11).

Therefore, without the additional condition of (4.2.11), it is possible to obtain different CEV limits under different normalizations. However, from the form of the limit in (4.3.4), it is clear that $\mu(\{\infty\} \times (y,\infty]) = G(\{\infty\})y^{-1}$. Thus, the exclusion of defective distributions in Theorem 4.3.1 is enough to avoid cases like the previous one.

Another subtlety highlighted in Example 4.3.6 is the idea that the degree of "extremeness" depends on our choice of normalization. In the first case, the mass settling at $\{\infty\}$ for G leads to mass on the line $\{\infty\} \times (0, \infty]$ for the CEV limit measure. This is due to the fact that a portion of the mass of the normalized joint distribution is pushed towards the x-axis as $t \to \infty$ in (4.3.8). This reflects the possibility that X can be asymptotically of higher order that Y (when $W = 0$), so Y would be considered "non-extreme" from the point of view of such X. On the other hand, normalizing X by $\alpha(t) = t^2$, X is asymptotically either of the same order as Y or of lower order (when $W = 1$), which is captured by the mass G assigns to $\{0\}$. The condition $\mu(\{\infty\} \times (y, \infty]) = 0$ requires that X be of asymptotic order no larger than Y.

We now present a case where the normalized kernel K does not have a unique limit.

Example 4.3.7. Suppose $Y \sim$ Pareto(1), and define X by

$$X = WY + (1 - W)2Y\,\mathbf{1}_{\{Y \in [0,\infty)\setminus\mathbb{N}\}}$$

where $W \sim$ Bernoulli(p) independent of Y. In other words, given $Y = y$, X takes the value y or $2y$ according to a coin flip, unless y is an integer, in which case X will be either y or 0.

Observe that (X, Y) follows a conditional model. Since $\mathsf{P}[Y \in \mathbb{N}] = 0$, we have

$$\begin{aligned}
\mathsf{P}[X \leq x, Y > y] &= \mathsf{P}[X \leq x,\, Y > y,\, Y \in [0,\infty)\setminus\mathbb{N}] \\
&= p\,\mathsf{P}(Y \leq x, Y > y) + (1-p)\,\mathsf{P}(2Y \leq x, Y > y) \\
&= p(y^{-1} - x^{-1})\,\mathbf{1}_{\{x \geq y\}} + (1-p)(y^{-1} - 2x^{-1})\,\mathbf{1}_{\{x \geq 2y\}},
\end{aligned}$$

and $t\,\mathsf{P}[X \le tx, Y > ty] = \mathsf{P}[X \le x, Y > y]$, which satisfies (4.2.9) and (4.2.10). However, the conditional distribution of X given Y is

$$K(y\,,\,\cdot) = \begin{cases} p\epsilon_y + (1-p)\epsilon_0 & y \in \mathbb{N} \\ p\epsilon_y + (1-p)\epsilon_{2y} & y \in [0,\infty)\backslash\mathbb{N} \end{cases},$$

so

$$K(t\,,\,t\,\cdot) = \begin{cases} p\epsilon_1 + (1-p)\epsilon_0 & t \in \mathbb{N} \\ p\epsilon_1 + (1-p)\epsilon_2 & t \in [0,\infty)\backslash\mathbb{N} \end{cases}.$$

$K(t, t\,\cdot)$ clearly does not converge, since we obtain different limits along the sequences $t_n = n$ and $t'_n = n/2$.

Example 4.3.7 draws attention to a technical difficulty that arises when considering conditional distributions $\mathsf{P}[X \in \cdot \,|\, Y = y]$. These are only specified up to sets of $\mathsf{P}[Y \in \cdot\,]$-measure zero. Hence, if Y is absolutely continuous, we can alter the conditional probability for a countable number of y without affecting the joint distribution.

Because the CEVM is formulated in terms of the joint distribution, it holds or fails to hold regardless of the version of conditional distribution we adopt. However, the convergence of the kernel as in (4.3.2) does depend on the version chosen. This is a mathematical weakness in the limit theory based on conditional distributions proposed by Heffernan and Tawn [45]. Nevertheless, in many common situations, the joint distribution of (X, Y) is described in terms of a natural choice of smooth version K, in which (4.3.2) may be used as a criterion to check whether a CEVM is appropriate. For example, X may arise as an explicit function of Y, in which case the representation (2.2.9) can be used. Also, if (X, Y) has a continuous density, then a natural choice of K is the one which is continuous in y, for which behaviour as in Example 4.3.7 is not possible.

4.4 General Normalization for X

Part of the usefulness of the CEVM stems from its ability to normalize X and Y differently, as in (4.2.8). However, the assumption $K \in D(G)$, phrased as in (4.3.2), imposes the same normalization for both. We now extend this condition to incorporate general linear normalizations of X, continuing to assume $F_Y \in D(G_1^*)$ (4.3.1).

We will assume the following generalization of (4.3.2): there exists a non-degenerate probability distribution G on $[-\infty, \infty)$ and normalization functions $\alpha > 0$, $\beta \in \mathbb{R}$, such that

$$K\big(t\,,\, [-\infty, \alpha(t)x + \beta(t)]\big) \Rightarrow G([-\infty, x]) \qquad \text{on } [-\infty, \infty]. \tag{4.4.1}$$

4.4.1 CEVM Properties

Consider the decomposition

$$t\,\mathsf{P}\left[\frac{X - \beta(t)}{\alpha(t)} \leq x, Y > ty\right] = \int_{(y,\infty]} t\,\mathsf{P}[Y \in tdu] K\big(tu\,,\, [-\infty, \alpha(t)x + \beta(t)]\big).$$

By Lemma A.0.2, this will converge provided $K\big(tu(t)\,,\, [-\infty, \alpha(t)x + \beta(t)]\big) \to \varphi_x(u)$ whenever $u(t) \to u > 0$. To understand the conditions under which (4.4.1) extends to this form, we follow a development similar to Section 2.2.1.

Given $\rho, k \in \mathbb{R}$, define the *generalized tail kernel* associated with a distribution G on $[-\infty, \infty]$ as the transition function $\kappa_G : (0, \infty) \times \mathcal{B}[-\infty, \infty] \to [0, 1]$ given by

$$\kappa_G(y\,,\, A) = G\big(y^{-\rho}[A - \psi(y)]\big), \tag{4.4.2}$$

where ψ is the function given in (4.2.2) (p. 93). Note that κ_G describes transitions between two different spaces. It is easy to see that ψ satisfies $\psi(uy) = u^\rho\psi(y) + \psi(u)$, implying that a kernel κ has the form (4.4.2) iff

$$\kappa(uy\,,\,A) = \kappa\big(y\,,\,u^{-\rho}[A-\psi(u)]\big).$$

Proposition 4.4.1. *Let $K : (0,\infty)\times\mathcal{B}[-\infty,\infty] \to [0,1]$ be a transition function satisfying (4.4.1), where G is non-degenerate. There exists a family of non-degenerate distributions $\{G_u : 0 < u < \infty\}$ on $[-\infty,\infty]$ such that*

$$K\big(tu\,,\,[-\infty,\alpha(t)x+\beta(t)]\big) \Rightarrow G_u([-\infty,x]) \quad \text{on } [-\infty,\infty], \quad 0 < u < \infty, \quad (4.4.3)$$

as $t\to\infty$ if and only if $\alpha,\beta\in \mathrm{ERV}_{\rho,k}$ as in (4.2.1) (p. 93). In this case, $G_1 = G$, and

$$K\big(tu_t\,,\,[-\infty,\alpha(t)x+\beta(t)]\big) \Rightarrow \kappa_G(u\,,\,[-\infty,x]) \qquad \text{on } [-\infty,\infty] \qquad (4.4.4)$$

whenever $u_t = u(t) \to u \in (0,\infty)$, i.e., the limit is a transition function of the form (4.4.2), where ρ,k are the ERV parameters of α,β.

Proof. Assume first that $\alpha,\beta\in\mathrm{ERV}_{\rho,k}$. Then for $u>0$,

$$K\big(tu\,,\,[-\infty,\alpha(t)x+\beta(t)]\big) = K\big(tu\,,\,\alpha(tu)\{h_t^{-1}(\cdot\,;u)[-\infty,x]\}+\beta(tu)\big),$$

where

$$h_t(y;u) = \frac{\alpha(tu)}{\alpha(t)}y + \frac{\beta(tu)-\beta(t)}{\alpha(t)}.$$

By (4.2.1), $h_t(y_t;u) \to h(y;u) = u^\rho y + \psi(u)$ whenever $y_t \to y\in\mathbb{R}$. Therefore, applying the second continuous mapping theorem (Lemma A.0.1, p. 132) to the weak convergence (4.4.1), we have

$$K\big(tu\,,\,[-\infty,\alpha(t)x+\beta(t)]\big) \Rightarrow (G\circ h^{-1}(\cdot\,;u))([-\infty,x]).$$

Hence, (4.4.3) holds with $G_u = \kappa_G(u, \cdot)$, and so $G_1 = G$. Furthermore, we have $h_t(x_t; u_t) \to h(x; u)$ whenever $u_t \to u > 0$, establishing (4.4.4).

For the converse, we employ convergence of types. Denote by H_t the distribution $K(t, \cdot)$. Then, on the one hand, we have $H_t([-\infty, \alpha(t)x + \beta(t)]) \Rightarrow G_1([-\infty, x])$. On the other hand, fixing $c > 0$, we have

$$H_t(\alpha(tc)x + \beta(tc)) = K\big((tc)c^{-1}\,,\,[-\infty,\,\alpha(tc)x + \beta(tc)]\big) \Rightarrow G_{c^{-1}}([-\infty, x]).$$

Convergence of types yields that $\alpha, \beta \in \mathrm{ERV}_{\rho,k}$, and

$$G_{c^{-1}}([-\infty, x]) = G_1([-\infty, c^{\rho}x + \psi(c)]),$$

with ψ as in (4.2.2). Using the identity (4.2.3) (p. 93), we find that G_u has the form (4.4.2), with $G = G_1$. $\square$

A consequence of Proposition 4.4.1 is that, in order to obtain a CEVM using (4.4.1), a necessary and sufficient condition is that α, β are ERV. Requiring G to be non-degenerate is necessary in order to apply convergence of types. For now, we continue to assume that Y is in the standardized domain of attraction.

Theorem 4.4.1. *Suppose (X, Y) is a random vector on $\mathbb{R} \times [0, \infty)$ with $F_Y \in D(G_1^*)$ (4.3.1), and $K(y, \cdot) = \mathsf{P}[X \in \cdot \,|\, Y = y]$ converges according to (4.4.1) for some normalizing functions $\alpha > 0$ and $\beta \in \mathbb{R}$ and non-degenerate limit distribution G on $[-\infty, \infty)$. Then, as $t \to \infty$,*

$$t\,\mathsf{P}\left[\left(\frac{X - \beta(t)}{\alpha(t)}, \frac{Y}{t}\right) \in \cdot\,\right] \xrightarrow{v} \mu(\cdot) \qquad \text{in } \mathbb{M}_+([-\infty, \infty] \times (0, \infty]) \tag{4.4.5}$$

where μ is a non-null Radon measure satisfying the conditional non-degeneracy conditions (4.2.11), if and only if $\alpha, \beta \in \mathrm{ERV}_{\rho,k}$. In this case, the limit measure μ is specified by

$$\mu\left([-\infty, x] \times (y, \infty]\right) = \int_{(y,\infty]} \nu_1(du)\,\mathsf{P}\left[\xi \le u^{-\rho}(x - \psi(u))\right], \quad x \in \mathbb{R},\ y > 0, \tag{4.4.6}$$

with ψ *as in* (4.2.2) *(p. 93) and* $\xi \sim G$. *The expression* (4.4.6) *is continuous in* x *and* y *if* $(\rho, k) \neq (0,0)$, *or if* G *is continuous.*

Proof. The convergence (4.4.5) to a limit μ satisfying (4.2.11) implies $\alpha, \beta \in$ ERV [44, Proposition 1]. Conversely, if $\alpha, \beta \in \mathrm{ERV}_{\rho,k}$, then the convergence (4.4.5) follows from Lemma A.0.4 (p. 133) in light of (4.4.4), yielding the limit in (4.4.6). We check that $\mu([-\infty, x] \times (y, \infty])$ is continuous when $(\rho, k) \neq (0,0)$. Indeed, applying dominated convergence, if $x_n \to x$, then

$$\mathsf{P}[\xi \leq u^{-\rho}(x_n - \psi(u))] \to \mathsf{P}[\xi \leq u^{-\rho}(x - \psi(u))]$$

for all except a countable number of u corresponding to discontinuities of the distribution function. Continuity in y is clear. Also, if $(\rho, k) = (0,0)$, then $\mu([-\infty, x] \times (y, \infty]) = y^{-1}G([-\infty, x])$, which is continuous if G is. In either case, $\mu([-\infty, x] \times (y, \infty])$ is non-degenerate in x because G is non-degenerate. Finally, $\mu(\{\infty\} \times (y, \infty]) = y^{-1}G(\{\infty\}) = 0$. Therefore, μ satisfies (4.2.11). $\square$

The limit measure is

$$\mu([-\infty, x] \times (y, \infty]) = \int_0^{y^{-1}} du\, \mathsf{P}\left[\xi \leq u^\rho x + \psi(u)\right], \tag{4.4.7}$$

where

$$u^\rho x + \psi(u) = \begin{cases} u^\rho(x + k\rho^{-1}) - k\rho^{-1} & \rho \neq 0 \\ x + k \log u & \rho = 0 \end{cases}.$$

Changing variables, we obtain the following expressions for μ according to (ρ, k):

$$\mu([-\infty, x] \times (y, \infty]) = \tag{4.4.8}$$
$$\begin{cases} \dfrac{1}{\rho|x+k\rho^{-1}|^{1/\rho}} \displaystyle\int_0^{|x+k\rho^{-1}|y^{-\rho}} u^{(1-\rho)/\rho}\, \mathsf{P}\left[\xi \le u \operatorname{sgn}(x+k\rho^{-1}) - k\rho^{-1}\right] du & \rho \neq 0 \\ \dfrac{1}{|k|e^{x/k}} \displaystyle\int_{-\infty}^{x \operatorname{sgn}(k) - |k| \log y} e^{u/|k|}\, \mathsf{P}\left[\xi \le u \operatorname{sgn}(k)\right] du & \rho = 0, \quad k \neq 0 \cdot \\ y^{-1}\, \mathsf{P}[\xi \le x] & \rho = 0, \quad k = 0 \end{cases}$$

Here $\operatorname{sgn}(v) = v/|v|\, \mathbf{1}_{\{v \neq 0\}}$, and we read the measure as $y^{-1}\,\mathsf{P}[\xi \le -k\rho^{-1}]$ when $x = -k\rho^{-1}$ for the case $\rho \neq 0$. Continuity in x and y when $(\rho, k) \neq (0,0)$ is apparent from the above expressions.

We now demonstrate a case where K satisfies (4.4.1), but (4.4.5) fails because α, β are not ERV.

Example 4.4.1. Consider $Y \sim$ Pareto(1), and $U \sim$ Uniform(0, 1), independent of Y. Put $X = Ue^Y$. Then

$$K(y\,,\, [0, x]) = \mathsf{P}[X \le x \,|\, Y = y] = \mathsf{P}[U \le xe^{-y}] = xe^{-y} \wedge 1.$$

In this case, polynomial scaling is not strong enough to give an informative limit, since

$$K(t\,,\, t^\rho[0, x]) = x^\rho t^\rho e^{-t} \wedge 1 \to 0.$$

The appropriate normalization would be exponential $\alpha(t) = e^t$:

$$K(t\,,\, \alpha(t)[0, x]) = xe^t e^{-t} \wedge 1 \to x \wedge 1 = G([0, x]).$$

In fact, by the convergence to types theorem, this the only normalization yielding a non-degenerate limit, up to asymptotic equivalence. However, since α is not regularly varying, Theorem 4.4.1 shows that (X, Y) cannot follow a CEVM. Indeed, consider

$$
\begin{aligned}
t\,\mathsf{P}[X \le \alpha(t)x, Y > ty] &= \int_{(y,\infty]} \nu_1(du) K(tu\,,\,[0, e^t x]) \\
&= \int_{((y\vee 1),\infty]} \nu_1(du)\{xe^{-t(u-1)} \wedge 1\} + \mathbf{1}_{\{y<1\}} \int_{(y,1]} \nu_1(du)\{xe^{-t(u-1)} \wedge 1\}.
\end{aligned}
$$

The first integral in the previous sum is bounded by $xy^{-1}e^{-t(y-1)} \to 0$. If $y \le 1$, the second integral approaches $\nu_1(y,1] = y^{-1} - 1$. Therefore, the limit is degenerate in x, violating conditional non-degeneracy (4.2.11).

In fact, we find that no choice of ERV normalization will lead to a CEVM. Indeed, suppose $\tilde{\alpha}, \tilde{\beta}$ are ERV. Then

$$
\begin{aligned}
t\,\mathsf{P}[X \le \tilde{\alpha}(t)x + \tilde{\beta}(t), Y > ty] &= \int_{(y,\infty]} \nu_1(du) K\big(tu\,,\,[0, \tilde{\alpha}(t)x + \tilde{\beta}(t)]\big) \\
&= \int_{(y,\infty]} \nu_1(du) \big\{e^{-tu}(\tilde{\alpha}(t)x + \tilde{\beta}(t)) \wedge 1\big\} \longrightarrow 0
\end{aligned}
$$

(see Section 4.2.1 (p. 93) for a summary of the asymptotic properties of ERV functions).

4.4.2 Standardization of X

Das and Resnick [25, Section 3.2] show that in certain cases, it is possible to standardize the X variable. Denote by x^* and x_* the upper and lower endpoints of the distribution of X respectively, i.e.,

$$
x^* = \sup\{x : F_X(x) < 1\} \qquad \text{and} \qquad x_* = \inf\{x : F_X(x) > 0\}.
$$

We will call $f : (0, \infty) \to (x_*, x^*)$ a *standardization function* if f is monotone and $\lim_{x\to\infty} f(x) \in \{x_*, x^*\}$. Following [25, Section 3], we will restrict attention to standardization by such functions. However, we have inverted the definition given in [25], in the sense that we will be using $f^{\leftarrow}$ to standardize rather than f.

For the purpose of this section, we extend the definition of $f^{\leftarrow}$ in order to invert right-continuous monotone functions which are either increasing or decreasing. Define

$$f^{\leftarrow}(x) = \begin{cases} \inf\{y : f(y) \geq x\} & \text{if } f \text{ is non-decreasing} \\ \inf\{y : f(y) \leq x\} & \text{if } f \text{ is non-increasing} \end{cases}.$$

Note that $f^{\leftarrow}$ is left-continuous for f non-decreasing and right-continuous for f non-increasing. The main property we shall be using is that

$$\begin{cases} f^{\leftarrow}(x) \leq y \iff x \leq f(y) & f \text{ non-decreasing} \\ f^{\leftarrow}(x) \leq y \iff x \geq f(y) & f \text{ non-increasing} \end{cases}.$$

The distinction between the two cases is a technicality which should not cause confusion in the following discussion. Also, we will say that a monotone function f has two "points of change" if there exist $x_1 < x_2 < x_3$ such that $f(x_1) < f(x_2) < f(x_3)$ for f non-decreasing, and with the opposite inequalities in the non-increasing case.

If (X, Y) satisfy (4.4.5) for some $\alpha > 0$ and β, then we say (X, Y) can be *standardized* if there exists a standardization function f such that

$$t\,\mathsf{P}\left[t^{-1}(f^{\leftarrow}(X), Y) \in \cdot\,\right] \xrightarrow{v} \mu^*(\cdot) \qquad \text{in } \mathbb{M}_+([0,\infty] \times (0,\infty]), \tag{4.4.9}$$

where μ^* is a non-null Radon measure. If the limit μ in (4.4.5) satisfies the conditional non-degeneracy conditions (4.2.11), then standardization is possible if and only if $(\rho, k) \neq (0, 0)$, i.e., μ is not a product. Because of the dependence on α and β, we can characterize functions f yielding (4.4.9) in the following way.

Proposition 4.4.2. *Suppose* (X, Y) *follow a CEVM, i.e.,* (4.4.5) *holds with* μ *satisfying the conditional non-degeneracy conditions* (4.2.11)*, and* $(\rho, k) \neq (0, 0)$*. Then a standardization function* f *standardizes* (X, Y) *in the sense of* (4.4.9), *where μ^* satisfies the conditional non-degeneracy conditions, if and only if*

$$\frac{f(tx)-\beta(t)}{\alpha(t)} \longrightarrow \varphi(x), \qquad x>0, \tag{4.4.10}$$

where φ has at least two points of change. In this case, μ and μ^* are related by

$$\mu^*([0,x]\times(y,\infty]) = \mu(A_\varphi(x)\times(y,\infty]),$$

where

$$A_\varphi(x) = \begin{cases} [-\infty,\varphi(x)] & f \text{ non-decreasing} \\ [\varphi(x),\infty] & f \text{ non-increasing} \end{cases}. \tag{4.4.11}$$

It follows that $\alpha, f \in \mathrm{ERV}$, although not necessarily with the same parameters as α,β. However, depending on the case, f can be expressed in terms of either β or α (see [24, Proposition 2.3.3]).

Proof. Suppose f is non-decreasing. Then for $x,y>0$, we can write

$$t\,\mathsf{P}\left[\frac{f^{\leftarrow}(X)}{t}\le x, \frac{Y}{t}>y\right] = t\,\mathsf{P}\left[\frac{X-\beta(t)}{\alpha(t)}\le \frac{f(tx)-\beta(t)}{\alpha(t)}, \frac{Y}{t}>y\right]. \tag{4.4.12}$$

If f satisfies (4.4.10), then (4.4.9) holds with

$$\mu^*([0,x]\times(y,\infty]) = \mu([-\infty,\varphi(x)]\times(y,\infty])$$

non-degenerate in x. On the other hand, if (4.4.9) holds, then (4.4.12) implies (4.4.10), and φ has at least two points of increase because μ^* is non-degenerate in x. The mass at $\{\infty\}$ condition in (4.2.11) follows from the fact that $\lim_{x\to\infty}\varphi(x)=\infty$ if f is non-decreasing (see (4.4.13) below). The case for f non-increasing is similar, after reversing the inequality for X on the right-hand side of (4.4.12). $\square$

Assuming (4.4.10) and $\alpha,\beta\in\mathrm{ERV}_{k,\rho}$, write

$$\frac{f(tx)-\beta(t)}{\alpha(t)} = \frac{\alpha(tx)}{\alpha(t)}\frac{f(tx)-\beta(tx)}{\alpha(tx)} + \frac{\beta(tx)-\beta(t)}{\alpha(t)}$$

and $c = \varphi(1)$. We find that φ has the form

$$\varphi(x) = \begin{cases} cx^\rho + k\rho^{-1}(x^\rho - 1) & \rho \neq 0 \\ c + k \log x & \rho = 0 \end{cases}. \tag{4.4.13}$$

That φ is non-constant imposes the constraint that $c \neq 0$ if $\rho \neq 0$, $k = 0$.

What if the conditional distribution of X given Y in fact satisfies the kernel convergence assumption (4.4.1)? We can then apply any standardization directly the conditional distribution via its transition function.

Proposition 4.4.3. *Suppose the transition function $K : (0, \infty) \times \mathcal{B}[-\infty, \infty] \to [0, 1]$ satisfies* (4.4.1) *for a probability distribution G on $[-\infty, \infty)$. If f is a standardization function satisfying* (4.4.10)*, then the transition function $K_f : (0, \infty) \times \mathcal{B}[0, \infty] \to [0, 1]$ defined as*

$$K_f(y\,,\, A) = K(y\,,\, f(A))$$

satisfies

$$K_f(t\,,\, t[0, x]) \Rightarrow G(A_\varphi(x)) =: G_f([0, x]) \qquad \text{on } [0, \infty],$$

with $A_\varphi(x)$ as in (4.4.11)*. Conversely, suppose $G([0, \infty)) = 1$, and*

$$K(t\,,\, t[0, x]) \Rightarrow G([0, x]) \qquad \text{on } [0, \infty]. \tag{4.4.14}$$

Then, given ERV functions $\alpha > 0$, $\beta \in \mathbb{R}$ on $(0, \infty)$, if f is a monotone function defined on $(0, \infty)$ satisfying (4.4.10)*, the transition function $\overline{K}_f : (0, \infty) \times \mathcal{B}[-\infty, \infty] \to [0, 1]$ given by*

$$\overline{K}_f(y\,,\, A) = K(y\,,\, f^{\leftarrow}(A))$$

satisfies

$$\overline{K}_f\big(t\,,\, [-\infty, \alpha(t)x + \beta(t)]\big) \Rightarrow G(A_{\varphi^\leftarrow}(x)) =: \overline{G}_f([-\infty, x]) \qquad \text{on } f([0, \infty]),$$

where

$$A_{\varphi^{\leftarrow}}(x) = \begin{cases} [0, \varphi^{\leftarrow}(x)] & f \text{ non-decreasing} \\ [\varphi^{\leftarrow}(x), \infty] & f \text{ non-increasing} \end{cases} .$$

Proof. Assume (4.4.1) first, and let f be a non-decreasing standardization function satisfying (4.4.10). Then,

$$\begin{aligned} K_f\big(t\,, t[0,x]\big) &= K\big(t\,,\,[-\infty, f(tx)]\big) \\ &= K\left(t\,,\,\alpha(t)\left[-\infty, \frac{f(tx)-\beta(t)}{\alpha(t)}\right] + \beta(t)\right) \Rightarrow G([-\infty, \varphi(x)]). \end{aligned}$$

On the other hand, if f satisfies (4.4.10) for $\alpha, \beta \in \mathrm{ERV}$, then inverting this relation yields

$$\frac{f^{\leftarrow}(\alpha(t)x+\beta(t))}{t} \longrightarrow \varphi^{\leftarrow}(x), \qquad x \in f((0,\infty)).$$

Consequently, for non-decreasing f,

$$\overline{K}_f\big(t\,,\,[-\infty, \alpha(t)x+\beta(t)]\big) = K\big(t\,,\,t[0, t^{-1}f^{\leftarrow}(\alpha(t)x+\beta(t))]\big) \Rightarrow G([0, \varphi^{\leftarrow}(x)]).$$

The case for non-increasing f is similar. □

Proposition 4.4.3 rephrases Das's result [24, Proposition 2.3.3] on the existence of standardization functions in terms of conditional distributions rather than joint distributions, Indeed, suppose K is a version of the conditional distribution $\mathsf{P}[X \in \cdot \,|\, Y = y]$, satisfying (4.4.1), where $\alpha, \beta \in \mathrm{ERV}_{\rho,k}$ with $(\rho, k) \neq (0,0)$. If F_Y is in the standardized domain of attraction, then (X,Y) follows a CEVM by Theorem 4.4.1. Furthermore, (X,Y) can be standardized in the sense of (4.4.9) [24, Proposition 2.3.3 (1)], and the standardization function f satisfies (4.4.10) by Proposition 4.4.2. This standard CEVM could equally have been obtained by applying Theorem 4.3.1 to the transition function K_f, a version of the conditional distribution $\mathsf{P}[f^{\leftarrow}(X) \in \cdot \,|\, Y = y]$, giving (4.4.9) directly. That (X,Y) follow a (non-standard) CEVM would then follow by unstandardizing the limit measure μ^*.

In Section 4.3.1 we discussed the properties of the limit measure μ in the standard case. In particular, if X belongs to the standardized domain of attraction, then G necessarily satisfies the moment condition $0 \le \mathsf{E}\,\xi \le 1$ (recall $\xi \sim G$). This is due to the fact that $\mu((1,\infty] \times (0,\infty])$ is bounded by x^{-1}. Using the standardization approach discussed above, we can derive necessary conditions on G when X belongs to a general domain of attraction.

Suppose there exist normalizing functions $c(t) > 0$ and $d(t)$ such that

$$t\,\mathsf{P}\left[\frac{X - d(t)}{c(t)} > x\right] \longrightarrow (1+\lambda x)^{-1/\lambda} \qquad x \in \mathbb{E}_\lambda, \tag{4.4.15}$$

implying that $c, d \in \mathrm{ERV}_{\lambda,1}$ (see Section 4.2.2, p. 94). Das and Resnick [25, Proposition 4.1] show that if (X, Y) follow a CEVM and (4.4.15) holds, then the vector (X, Y) belongs to a multivariate domain of attraction provided $\lim_{t\to\infty} \alpha(t)/c(t) \in [0,\infty)$.

For simplicity, we consider the case where the conditional distribution of X given Y satisfies

$$K\big(t\,,\,[-\infty, c(t)x + d(t)]\big) \Rightarrow G([-\infty, x]),$$

i.e., (4.4.1) holds under the same normalization as in (4.4.15). Then d is a standardization function satisfying (4.4.10), and

$$\varphi(x) = \begin{cases} \lambda^{-1}(x^\lambda - 1) & \lambda \ne 0 \\ \log x & \lambda = 0 \end{cases}$$

(see (4.2.7), p. 95). Theorem 4.3.1 gives a standard CEVM for $(d^\leftarrow(X), Y)$, and furthermore, $d^\leftarrow(X)$ belongs to the standardized domain of attraction. Therefore, the distribution G must satisfy

$$\int_0^\infty \mathsf{P}[\xi > \varphi(x)]dx \leq 1.$$

Depending on λ, this reduces to

$$\begin{cases} \mathsf{E}\,\xi^{1/\lambda}\,\mathbf{1}_{\{\xi>0\}} \leq \lambda^{-1/\lambda} & \lambda > 0 \\ \mathsf{E}(-1/\xi)^{1/|\lambda|}\,\mathbf{1}_{\{\xi<0\}} \leq |\lambda|^{1/|\lambda|} & \lambda < 0 \\ \mathsf{E}\,e^{\xi} \leq 1 & \lambda = 0 \end{cases} \cdot$$

Thus, we obtain a different condition for each class of extreme value distribution. In the Fréchet case, we have a bound on the $1/\lambda$-th moment of the right tail. If the domain of attraction is Weibull, this becomes an integrability condition near 0. Finally, in the Gumbel case, the right tail of ξ is exponentially bounded, so all right-tail moments exist.

4.4.3 Relation to the Heffernan and Tawn Model

The CEVM of Theorem 4.4.1 is inspired by the statistical model proposed by Heffernan and Tawn [45]. We now discuss some links between this work and the CEVM.

Where Heffernan and Tawn's model is based on the convergence of conditional distributions, as in (4.4.1), the CEVM focuses on limits of joint distributions. Theorem 4.4.1 shows that Heffernan and Tawn's assumption [45, Equation (3.1)] leads to a CEVM provided the normalization functions α and β are ERV. The fact that the convergence (4.4.1) is required to hold at all points x suggests that they are expecting a continuous limit. Instead, we have framed the assumption in the more theoretically appealing context of weak convergence. Also, Heffernan and Tawn standardize the conditioning variable to a Gumbel domain of attraction rather than Fréchet, which is our condition (4.3.1), but this is a minor point.

Also, Example 4.3.7 demonstrates a theoretical disadvantage to working with conditional distributions. A condition such as (4.4.1) is tacitly, if not explicitly, assuming a particular version of the conditional distribution. This issue cannot be ignored, since Example 4.3.7 shows that (4.4.1) holding for one particular version does not imply that it holds for every version. This question of version is not addressed by Heffernan and Tawn. However, it should not pose a problem if we assume that the distributions are absolutely continuous, as is common in statistical contexts.

Another interesting point concerns the normalization functions α and β. If a non-degenerate CEVM holds, then these are necessarily ERV. It is not clear whether Heffernan and Tawn recognized this as a theoretical result, but they do assume a parametric form for these functions which is very similar to ERV. They specify

$$\alpha(y) = b_{|i}(y) := y^{b_{|i}} = y^{\rho}$$

for some constant $\rho < 1$ and

$$\beta(y) = a_{|i}(y) := \begin{cases} ay & 0 \le \rho < 1, \quad \text{with } a \in [0,1] \\ c - d\log y & \rho < 0 \quad \text{with } a = 0,\, c \in \mathbb{R},\, d \in [0,1] \end{cases}.$$

Although more general models are possible, the form of the ERV limit function ψ in (4.2.2) (p. 93) suggests that a parametric approach is indeed reasonable.

4.5 General Normalizations for both X and Y

Up until now, we have been assuming that Y belongs to the standardized Fréchet domain of attraction: $t\,\mathsf{P}[Y > ty] \to y^{-1}$ for $y > 0$. We wish to extend the result of Theorem 4.4.1 to the case where Y belongs to a general domain of attraction:

$$t\,\mathsf{P}[Y > a(t)y + b(t)] \longrightarrow (1+\gamma y)^{-1/\gamma} \qquad y \in \mathbb{E}_\gamma, \tag{4.5.1}$$

where $\mathbb{E}_\gamma := \{y : 1 + \gamma y > 0\}$. See Section 4.2.2 (p. 94) for further details on domains of attraction. We will assume that $b(t)$ is given by (4.2.6).

An important consideration in the previous development is that convergence results depend on properties of the particular choice of version $K(y,\cdot)$ of the conditional distribution $\mathsf{P}[X \in \cdot \,|\, Y = y]$. Because Y is now normalized according to a and b, the condition (4.4.1) may no longer be sufficient to obtain a general CEVM limit, as in (4.2.8) (p. 95).

If it were known that $K\big(a(t)u + b(t)\,,\, [-\infty, \alpha(t)x + \beta(t)]\big) \to \varphi_x(u)$ for $u > 0$, then (4.2.8) should follow from arguments similar to those in Section 4.4.1. On the other hand, Heffernan and Resnick [44] argue that (4.2.8) reduces to (4.4.5) by standardizing Y using the transformation $Y \mapsto b^{\leftarrow}(Y)$. Hence, if K^*, a specific version of $\mathsf{P}[X \in \cdot \,|\, b^{\leftarrow}(Y) = y]$, satisfies (4.4.1), then $(X, b^{\leftarrow}(Y))$ follows a CEVM under approriate normalization of X by Theorem 4.4.1, and (4.2.8) should follow from (4.4.5) by untransforming. We now examine the consistency of these two approaches.

4.5.1 Kernel Asymptotics

The transition function $K : (-\infty, \infty) \times \mathcal{B}[-\infty, \infty] \to [0, 1]$ will continue to denote a specific version of the conditional distribution of X given Y, i.e., ,

$$K(y\,,\,\cdot) = \mathsf{P}[X \in \cdot \,|\, Y = y].$$

Moving towards the transformation approach described above, we first argue that we can express a version of the conditional distribution of X given $b^{\leftarrow}(Y)$ in terms of K.

First, recall that the convergence (4.5.1), where b is given by (4.2.6), implies that $a, b \in \mathrm{ERV}_{\gamma,1}$. Hence, $a \in \mathrm{RV}_{\gamma}$, and

$$\frac{b(tx) - b(t)}{a(t)} \longrightarrow \begin{cases} \dfrac{x^{\gamma} - 1}{\gamma} & \gamma \neq 0 \\ \log x & \gamma = 0 \end{cases}, \qquad x > 0. \tag{4.5.2}$$

Inverting (4.5.2) gives

$$\frac{b^{\leftarrow}(a(t)x + b(t))}{t} \longrightarrow \begin{cases} (1 + \gamma x)^{1/\gamma} & \gamma \neq 0 \\ e^{x} & \gamma = 0 \end{cases}, \qquad x \in \mathbb{E}_{\gamma}. \tag{4.5.3}$$

Furthermore, recall that if $\tilde{b}$ is any function on $(0, \infty)$ satisfying

$$\frac{\tilde{b}(t) - b(t)}{a(t)} \longrightarrow 0 \qquad \text{as } t \to \infty, \tag{4.5.4}$$

then (4.5.1), (4.5.2), and (4.5.3) hold with b replaced by $\tilde{b}$. We now verify that we can choose such a $\tilde{b}$ which is invertible.

Lemma 4.5.1. *There exists a function b^* satisfying* (4.5.4) *that is continuous and strictly monotone.*

Proof. We consider cases on γ. If $\gamma = 0$, then $b \in \Pi(a)$. Then we can find $\bar{b}$ continuous, strictly increasing such that $(\bar{b}(t) - b(t))/a(t) \to 1$ by [76, Proposition 0.16]. The choice $b^*(x) = \bar{b}(e^{-1}x)$ satisfies (4.5.4). Otherwise, suppose $\gamma > 0$. Then $b \in \mathrm{RV}_\gamma$, and $b(t)/a(t) \to \gamma^{-1}$ [33, Theorem B.2.2 (1)]. Consequently, [77, Proposition 2.6 (vii)] gives a continuous, strictly increasing function $b^* \sim b$. Writing

$$\frac{b^*(t) - b(t)}{a(t)} = \frac{b(t)}{a(t)} \left[\frac{b^*(t)}{b(t)} - 1 \right]$$

shows that b^* satisfies (4.5.4). Finally, if $\gamma < 0$, then $b(\infty) = \lim_{t\to\infty} b(t)$ exists finite, $b(\infty) - b \in \mathrm{RV}_\gamma$, and $(b(\infty) - b(t))/a(t) \to -\gamma^{-1}$. Choose $\bar{b}$ continuous, strictly decreasing, with $\bar{b} \sim (b(\infty) - b)$, and set $b^* = b(\infty) - \bar{b}$. □

Henceforth, b^* will denote a continuous, strictly monotone function satisfying (4.5.4). The advantage to working with b^* is that $b^{*\leftarrow}(b^*(x)) = b^*(b^{*\leftarrow}(x)) = x$. By (4.5.2), $Y^* = b^{*\leftarrow}(Y)$ belongs to the standard domain of attraction when (4.5.1) holds:

$$t\,\mathsf{P}[Y^* > ty] = t\,\mathsf{P}\left[\frac{Y - b^*(t)}{a(t)} > \frac{b^*(ty) - b^*(t)}{a(t)}\right] \longrightarrow y^{-1}, \qquad y > 0.$$

We argue that when $K(y, \cdot) = \mathsf{P}[X \in \cdot \,|\, Y = y]$, the transition function

$$K^*(y\,,\,\cdot) := K(b^*(y)\,,\,\cdot) \tag{4.5.5}$$

is a version of the conditional distribution $\mathsf{P}[X \in \cdot \,|\, Y^* = y]$.

Proposition 4.5.1. *For measurable A and $y > 0$, we have*

$$\mathsf{P}[X \in A, Y^* > y] = \int_{(y,\infty)} K(b^*(u)\,,\,A)\,\mathsf{P}[Y^* \in du]. \tag{4.5.6}$$

Proof. Write

$$\mathsf{P}[X \in A, Y^* > y] = \mathsf{P}[X \in A, Y > b^*(y)] = \int_{(b^*(y),\infty)} K(u\,,\, A)\,\mathsf{P}[Y \in du]$$
$$= \int_{(b^*(y),\infty)} K\big(b^*(b^{*\leftarrow}(u))\,,\, A\big)\,\mathsf{P}[Y \in du],$$

using the fact that $b^*(b^{*\leftarrow}(u)) = u$ for all u, and change variables according to the transformation $T = b^{*\leftarrow}$. Since $T^{-1}(y,\infty) = \{x : b^{*\leftarrow}(x) > y\} = (b^*(y),\infty)$, the result follows. □

Note that (4.5.6) is not necessarily true for the function b given by (4.2.6), unless $\mathsf{P}[b(b^{\leftarrow}(Y)) \neq Y] = 0$.

Next, we show that the two approaches to the CEVM discussed at the beginning of Section 4.5, the direct approach and the standardization approach, are indeed consistent. That K^* converges to a family of distributions under scaling of the initial state, in the sense of (4.4.3) (p. 109), is equivalent to the same for K with initial state normalized by a and b.

Proposition 4.5.2. *Suppose Y is a random variable with distribution satisfying (4.5.1), and let K^* be given by (4.5.5). Given normalization functions $\alpha(t) > 0$ and $\beta(t) \in \mathbb{R}$, there exists a transition function $\phi^* : (0,\infty) \times \mathcal{B}[-\infty,\infty] \to [0,1]$ such that, as $t \to \infty$,*

$$K^*\big(tu_t\,,\, [-\infty, \alpha(t)x + \beta(t)]\big) \Rightarrow \phi^*(u\,,\, [-\infty, x]) \qquad \text{on } [-\infty,\infty] \tag{4.5.7}$$

whenever $u_t \to u \in (0,\infty)$, if and only if there exists a transition function $\phi : \mathbb{E}_\gamma \times \mathcal{B}[-\infty,\infty] \to [0,1]$ such that, as $t \to \infty$,

$$K\big(a(t)u_t + b(t)\,,\, [-\infty, \alpha(t)x + \beta(t)]\big) \Rightarrow \phi(u\,,\, [-\infty, x]) \qquad \text{on } [-\infty,\infty] \tag{4.5.8}$$

whenever $u_t \to u \in \mathbb{E}_\gamma$. If these convergences hold, then *(i)* $\alpha, \beta \in \mathrm{ERV}$;

(ii) $\phi^* = \kappa_{G^*}$, *a generalized tail kernel* (4.4.2) *with* $G^* = \phi^*(1, \cdot)$;

(iii) $\phi(u\,, A) = \kappa_G((1+\gamma u)^{1/\gamma}\,, A)$, *where* κ_G *is a generalized tail kernel with* $G = \phi(0, \cdot)$; *and*

(iv) *the two transition functions are related by* $G = G^*$.

Proof. Abbreviate $a_t = a(t)$ and $b_t = b(t)$. The convergences (4.5.2) and (4.5.3) are in fact locally uniform on $(0, \infty)$ (see Section 4.2.1, p. 93). Since b^* satisfies (4.5.4), it follows that

$$\frac{b^*(tu_t) - b_t}{a_t} \longrightarrow \frac{u^\gamma - 1}{\gamma} \qquad \text{whenever } u_t \to u \in (0, \infty),$$

and

$$\frac{b^{*\leftarrow}(a_t u_t + b_t)}{t} \longrightarrow (1+\gamma u)^{1/\gamma} \qquad \text{whenever } u_t \to u \in \mathbb{E}_\gamma.$$

Assuming (4.5.7), for $u_t \to u \in \mathbb{E}_\gamma$ we have

$$\begin{aligned} &K\big(a(t)u_t + b(t)\,, [-\infty, \alpha(t)x + \beta(t)]\big) \\ &\qquad = K\Big(b^*\big(t\{t^{-1}b^{*\leftarrow}(a_t u_t + b_t)\}\big)\,, [-\infty, \alpha(t)x + \beta(t)]\Big) \\ &\qquad = K^*\big(t\{t^{-1}b^{*\leftarrow}(a_t u_t + b_t)\}\,, [-\infty, \alpha(t)x + \beta(t)]\big) \\ &\qquad \Rightarrow \phi^*\big((1+\gamma u)^{1/\gamma}\,, [-\infty, x]\big) =: \phi(u\,, [-\infty, x]) \end{aligned}$$

Conversely, if (4.5.8) holds, then for $u_t \to u > 0$,

$$\begin{aligned} &K^*\big(tu_t\,, [-\infty, \alpha(t)x + \beta(t)]\big) \\ &\qquad = K\big(a_t \cdot a_t^{-1}(b^*(tu_t) - b_t) + b_t\,, [-\infty, \alpha(t)x + \beta(t)]\big) \\ &\qquad \Rightarrow \phi\big(\gamma^{-1}(u^\gamma - 1)\,, [-\infty, x]\big) =: \phi^*(u\,, [-\infty, x]) \end{aligned}$$

In either case, $G := \phi(0, \cdot) = \phi^*(1, \cdot) =: G^*$. Proposition 4.4.1 shows that α and β are ERV and $\phi^* = \kappa_{G^*}$. Consequently, $\phi(u\,, \cdot) = \kappa_G((1+\gamma u)^{1/\gamma}\,, \cdot)$. $\square$

Therefore, by Proposition 4.4.1 (p. 109), if there exists a non-degenerate distribution G on $[-\infty, \infty)$ such that

$$K^*\big(t\,,\,[-\infty, \alpha(t)x+\beta(t)]\big) = K\big(b^*(t)\,,\,[-\infty, \alpha(t)x+\beta(t)]\big) \Rightarrow G([-\infty, x]) \quad (4.5.9)$$

with $\alpha, \beta \in \mathrm{ERV}$, then (4.5.8) holds.

How can we apply Proposition 4.5.2 starting from an assumption like (4.5.9) on the kernel K rather than K^*? Because $b^*(b^{*\leftarrow}(t)) = t$, (4.5.9) can be written as

$$K\big(t\,,\,[-\infty, \alpha \circ b^{*\leftarrow}(t)x + \beta \circ b^{*\leftarrow}(t)]\big) \Rightarrow G([-\infty, x]) \qquad \text{as } t \to y^*,$$

where y^* denotes the upper endpoint of the distribution of Y, written as $y^* = \sup\{y : F_Y(y) < 1\}$. Therefore, we require there to exist a non-degenerate distribution G and normalization functions $\tilde{\alpha} > 0$ and $\tilde{\beta}$ such that

$$K\big(t\,,\,[-\infty, \tilde{\alpha}(t)x + \tilde{\beta}(t)]\big) \Rightarrow G([-\infty, x]) \qquad \text{as } t \to y^*, \qquad (4.5.10)$$

and $\alpha = \tilde{\alpha} \circ b^*$, $\beta = \tilde{\beta} \circ b^* \in \mathrm{ERV}$.

4.5.2 CEVM Properties

Using the standardization approach discussed in the previous section, we obtain a CEVM when Y belongs to a general domain of attraction.

Theorem 4.5.1. *Suppose (X, Y) is a random vector on $\mathbb{R}^2$, where $F_Y \in D(G_\gamma)$* (4.5.1)*, and $K(y, \cdot) = \mathsf{P}[X \in \cdot \,|\, Y = y]$ converges according to* (4.5.10)*, for some normalizing functions $\tilde{\alpha} > 0$ and $\tilde{\beta} \in \mathbb{R}$ and non-degenerate limit distribution G on $[-\infty, \infty)$. Let b^* be the function satisfying* (4.5.4) *given by Lemma 4.5.1, and put $\alpha = \tilde{\alpha} \circ b^*$, $\beta = \tilde{\beta} \circ b^*$. Then, as $t \to \infty$,*

$$t\,\mathsf{P}\left[\left(\frac{X - \beta(t)}{\alpha(t)}, \frac{Y - b(t)}{a(t)}\right) \in \cdot\,\right] \xrightarrow{v} \mu(\cdot) \qquad \text{in } \mathbb{M}_+([-\infty, \infty] \times \overline{\mathbb{E}}_\gamma), \quad (4.5.11)$$

where μ is a non-null Radon measure satisfying the conditional non-degeneracy conditions (4.2.11), *if and only if $\alpha, \beta \in \mathrm{ERV}_{\rho,k}$. In this case, the limit measure μ is specified by*

$$\mu([-\infty, x] \times (y, \infty]) = \int_0^{(1+\gamma y)^{-1/\gamma}} du\, \mathsf{P}\left[\xi \le u^\rho x + \psi(u)\right], \qquad x \in \mathbb{R},\ y \in \mathbb{E}_\gamma, \tag{4.5.12}$$

with ψ as in (4.2.2) *(p. 93). The expression* (4.5.12) *is continuous in x and y if* $(\rho, k) \neq (0, 0)$.

Proof. First, observe that $Y^* = b^{*\leftarrow}(Y) \in D(G_1^*)$. Defining the transition function $K^*(y, \cdot) = \mathsf{P}[X \in \cdot \,|\, Y^* = y]$ as in (4.5.5), our hypotheses imply (4.5.9). Therefore, if $\alpha, \beta \in \mathrm{ERV}_{\rho,k}$, then by Theorem 4.4.1, we have

$$t\,\mathsf{P}\left[\left(\frac{X - \beta(t)}{\alpha(t)}, \frac{Y^*}{t}\right) \in \cdot\right] \xrightarrow{v} \mu^*(\cdot) \qquad \text{in } \mathbb{M}_+([-\infty, \infty] \times (0, \infty]),$$

where μ^* is defined by

$$\mu^*([-\infty, x] \times (y, \infty]) = \int_0^{y^{-1}} du\, \mathsf{P}\left[\xi \le u^\rho x + \psi(u)\right], \quad x \in \mathbb{R},\ y > 0,$$

conditionally non-degenerate. Consequently, for $x \in \mathbb{R}$ and $y \in \mathbb{E}_\gamma$,

$$\begin{aligned} t\,\mathsf{P}&\left[\frac{X - \beta(t)}{\alpha(t)} \le x, \frac{Y - b(t)}{a(t)} > y\right] \\ &= t\,\mathsf{P}\left[\frac{X - \beta(t)}{\alpha(t)} \le x, \frac{Y^*}{t} > \frac{b^{*\leftarrow}(a(t)y + b(t))}{t}\right] \\ &= \int_0^{(1+\gamma y)^{-1/\gamma}} du\, \mathsf{P}\left[\xi \le u^\rho x + \psi(u)\right] = \mu([-\infty, x] \times (y, \infty]), \end{aligned}$$

and the marginal transformation of Y does not affect conditional non-degeneracy or continuity. Conversely, (4.5.11) implies that $\alpha, \beta \in \mathrm{ERV}$ [44, Proposition 1]. □

Remark. Instead of standardizing Y, we could equally have used the convergence (4.5.8), which holds under our assumptions by Propositions 4.4.1 and 4.5.2.

Recalling the forms of the limit measure given in Section 4.4.1, we can express the limit measure in (4.5.12) as

$$\mu([-\infty, x] \times (y, \infty]) =$$
$$\begin{cases} \dfrac{1}{\rho |x + k\rho^{-1}|^{1/\rho}} \displaystyle\int_0^{|x+k\rho^{-1}|(1+\gamma y)^{-\rho/\gamma}} u^{(1-\rho)/\rho} \, \mathsf{P}\left[\xi \le u \operatorname{sgn}(x + k\rho^{-1}) - k\rho^{-1}\right] du & \\ & \rho \neq 0 \\ \dfrac{1}{|k| e^{x/k}} \displaystyle\int_{-\infty}^{x \operatorname{sgn}(k) - |k|\gamma^{-1}\log(1+\gamma y)} e^{u/|k|} \, \mathsf{P}\left[\xi \le u \operatorname{sgn}(k)\right] du & \rho = 0, \ k \neq 0 \\ (1+\gamma y)^{-1/\gamma} \, \mathsf{P}[\xi \le x] & \rho = 0, \ k = 0 \end{cases}$$

where $\operatorname{sgn}(v) = v/|v| \, \mathbf{1}_{\{v \neq 0\}}$, and we read the measure as $(1+\gamma y)^{-1/\gamma} \, \mathsf{P}[\xi \le -k\rho^{-1}]$ when $x = -k\rho^{-1}$ for the case $\rho \neq 0$.

In Example 4.4.1 (p. 112), we presented a transition function satisfying (4.4.1) which did not lead to a CEVM when paired with $Y \in D(G_1^*)$. We now show that a non-degenerate CEVM may be obtained if Y belongs to a non-standardized domain of attraction.

Example 4.5.1. Consider $Y \sim \text{Exp}(1)$, and $U \sim \text{Uniform}(0, 1)$, independent of Y. Put $X = Ue^Y$. Note that $Y \in D(G_0)$ with $a(t) \equiv 1$, $b(t) = \log t$, since for $y \in \mathbb{R}$,

$$t \, \mathsf{P}(Y > y + \log t) = te^{-y-\log t} = e^{-y}.$$

A version of the conditional distribution is given by

$$K(y\,,\,[0, x]) = \mathsf{P}[X \le x \,|\, Y = y] = \mathsf{P}[U \le xe^{-y}] = xe^{-y} \wedge 1.$$

Taking $\tilde{\alpha}(t) = e^t$, we saw in Example 4.4.1 that

$$K(t\,,\,\tilde{\alpha}(t)[0, x]) \Rightarrow x \wedge 1 = G([0, x]),$$

although $\tilde{\alpha}$ is not regularly varying. Since b is continuous and strictly monotone, set $\alpha(t) = \tilde{\alpha}(b(t)) = t$. Then

$$K^*(t\,,\,t[0,x]) = K(b(t)\,,\,\tilde{\alpha}(b(t))[0,x]) \Rightarrow G([0,x]),$$

and $\alpha(t) \in \mathrm{RV}_1$. Hence, $K^*(tu\,,\,\alpha(t)[0,x]) \Rightarrow xu^{-1} \wedge 1 = G(u^{-1}[0,x])$, and $K^*(y,\cdot) = \mathsf{P}[X \in \cdot \,|\, e^Y = y]$. On the other hand, note that for $u \in \mathbb{R}$,

$$K\big(a(t)u + b(t)\,,\,\alpha(t)[0,x]\big) = txe^{-u-\log t} \wedge 1 = xe^{-u} \wedge 1 = G\big((e^u)^{-1}[0,x]\big).$$

This illustrates the equivalence presented in Proposition 4.5.2 (p. 124). Now, for $x > 0$, $y > 0$, the joint distribution is given by

$$\mathsf{P}[X \le x, Y > y] = \int_{\log x \vee y}^{\infty} xe^{-2u}du + \int_{y}^{\log x} e^{-u}du\,\mathbf{1}_{\{y<\log x\}}\,.$$

Therefore, for $x > 0$, $y \in \mathbb{R}$, and large t, we have

$$t\,\mathsf{P}[X \le tx, Y > y + \log t] = \left\{ \begin{array}{ll} \dfrac{xe^{-2y}}{2} & \text{if } \log x \le y \\ e^{-y} - \dfrac{1}{2x} & \text{if } \log x > y \end{array} \right\} = \mu([0,x] \times (y,\infty]),$$

and (X,Y) follow a CEVM by Theorem 4.5.1.

4.6 Conclusions and Future Directions

Although dealing with conditional distributions in general raises certain issues surrounding identifiability, in many statistical contexts a conditional formulation such as (4.4.1) is convenient. For example, it may be appropriate to model X as an explicit function of Y. Also, if we are working with distributions that have continuous densities, the natural choice of version of the conditional distribution is the absolutely continuous one, and other simplifications may be afforded.

The above development suggests that in such cases, the approach of Heffernan and Tawn [45] is reasonable, and will generally lead to a fairly parsimonious extremal model which can account for varying degrees of asymptotic independence. Heffernan and Tawn propose a semiparametric model, where the limit distribution G is estimated nonparametrically, and the normalization functions α and β belong to a parametric family. The extended regular variation of α and β provides some justification for this last assumption. Furthermore, the formulas for the limit measure derived above show that by modeling conditional distributions, we obtain a simpler CEV model parametrized by the distribution G and the pair (ρ, k), along with γ, the extreme value index of Y.

The question of fitting a bivariate CEV model has been considered by Das and Resnick [26] and by Fougères and Soulier [38]. These authors discuss statistics for detecting a CEV model and estimating the normalizing functions. However, many open questions remain, such as the asymptotic distributions of such estimators, and the appropriate method for nonparametric estimation of G. These problems may presumably be simplified substantially through the use of standardization for both X and Y.

Also, a natural extension of the bivariate model discussed above would be to consider the conditional formulation for higher-dimensional vectors. Indeed, this was the original intention of Heffernan and Tawn, who apply their methodology to a five-dimensional air pollution dataset. It is not clear what would be the appropriate formulation of a model conditioning on more than one extreme variable, nor the connections between such a model and the usual multivariate domain of attraction. In particular, cases where asymptotic independence is present between some pairs of variables but not others would require careful treatment. However, a model based on conditional distributions as developed above should prove useful.

APPENDIX A

TECHNICAL LEMMAS

Here we collect lemmas needed to prove convergence of integrals of the form $\int f_n \, d\mu_n$, assuming that $f_n \to f$ and $\mu_n \to \mu$ in their respective spaces. An example is the *Second Continuous Mapping Theorem* [12, Theorem 5.5, p. 34]:

Lemma A.0.1. *Assume $\mathbb{E}$ and $\mathbb{E}'$ are complete separable (cs) metric spaces, and for $n \geq 0$, $h_n : \mathbb{E} \to \mathbb{E}'$ are measurable. Define*

$$A = \{x \in \mathbb{E} : h_n(x_n) \to h_0(x) \text{ whenever } x_n \to x\}.$$

If P_n, $n \geq 0$ are probability measures on $\mathbb{E}$ with $P_n \Rightarrow P_0$, and $h_n \to h_0$ almost uniformly, in the sense that $P(A) = 1$, then $P_n \circ h_n^{-1} \Rightarrow P_0 \circ h_0^{-1}$ in $\mathbb{E}'$.

This result provides a way to handle the convergence of a family of integrals:

Lemma A.0.2. *In addition to the assumptions of Lemma A.0.1, require $\mathbb{E}' = \mathbb{R}$ and $\{h_n, n \geq 0\}$ is uniformly bounded, so that $\sup_{n\geq 0} \sup_{x\in\mathbb{E}} |h_n(x)| < \infty$.*

(a) *As $n \to \infty$,*

$$\int_{\mathbb{E}} h_n \, dP_n \longrightarrow \int_{\mathbb{E}} h_0 \, dP_0.$$

(b) *Suppose additionally that $\mathbb{E}$ is locally compact with a countable base (lccb), and $\mu_n \stackrel{v}{\to} \mu_0$ in $\mathbb{M}_+(\mathbb{E})$ with $\mu_0(A^c) = 0$. If there exists a compact set $B \in \mathcal{K}(\mathbb{E})$ with $\mu_0(\partial B) = 0$ such that $h_n(x) = 0$, $n \geq 0$ whenever $x \notin B$ (i.e., B is a common compact support of the h_n), then*

$$\int_{\mathbb{E}} h_n \, d\mu_n \longrightarrow \int_{\mathbb{E}} h_0 \, d\mu_0.$$

Proof. (a) If $X_n \sim P_n$ for $n \geq 0$, then $h_n(X_n) \Rightarrow h_0(X_0)$. The uniform boundedness of the h_n guarantees that $\mathsf{E}\, h_n(X_n) \to \mathsf{E}\, h_0(X_0)$.

(b) View B as a compact subspace of $\mathbb{E}$ inheriting the relative topology. Then, assuming $\mu(B) > 0$ to rule out a trivial case, define probabilities on B by $P_n(\cdot) = \mu_n(\,\cdot \cap B)/\mu_n(B)$, $n \geq 0$. Since $\mu_n(\,\cdot \cap B) \xrightarrow{v} \mu_0(\,\cdot \cap B)$ by [37, Proposition 3.3], and B is compact, we get $P_n \Rightarrow P_0$. Denote by h'_n, $n \geq 0$, the restrictions of the h_n to B. Observe that for any $x \in A \cap B$, we have $h'_n(x_n) \to h'(x)$ whenever $x_n \to x$ in B, and $P(A^c \cap B) \leq \mu(A^c)/\mu(B) = 0$. Therefore, we can apply part (a) to get

$$\int_{\mathbb{E}} h_n \, d\mu_n = \int_{\mathbb{E}} h_n \, \mathbf{1}_B \; d\mu_n = \mu_n(B) \int_B h'_n \, dP_n$$
$$\longrightarrow \mu_0(B) \int_B h'_0 \, dP_0 = \int_{\mathbb{E}} h_0 \, d\mu_0. \qquad \square$$

A convenient specialization of Lemma A.0.2 (b) is the following.

Lemma A.0.3. *Suppose* $\mathbb{E}$ *is lccb and* $\mu_n \xrightarrow{v} \mu$ *in* $\mathbb{M}_+(\mathbb{E})$*. If* $f : \mathbb{E} \to \mathbb{R}$ *is continuous and bounded, and* $B \in \mathbb{E}$ *is relatively compact with* $\mu(\partial B) = 0$*, then*

$$\int_B f \, d\mu_n \longrightarrow \int_B f \, d\mu.$$

Take $h_n = f\,\mathbf{1}_B$ for $n \geq 0$. Since $f\,\mathbf{1}_B$ is continuous except possibly on ∂B, we have $\mu(A^c) \leq \mu(\partial B) = 0$.

The next result is used to extend convergence of substochastic transition functions to vague convergence of measures on a larger space.

Lemma A.0.4. *Let* $\mathbb{E} \subset [0,\infty]^m$ *and* $\mathbb{E}' \subset [0,\infty]^{m'}$ *be two lccb spaces. Suppose that* $\{p^{(t)}, t \geq 0\}$*, are substochastic transition functions on* $\mathbb{E} \times \mathcal{B}(\mathbb{E}')$*. This means* $p^{(t)}(\cdot, B)$ *is a measurable function for any fixed* $B \in \mathcal{B}(\mathbb{E}')$*,* $p^{(t)}(x, \cdot)$ *is a measure for any* $x \in \mathbb{E}$*, and* $\sup_{t\geq 0} \sup_{u\in\mathbb{E}} p^{(t)}(u, \mathbb{E}') \leq 1$*. Assume there is a set* $A \subset \mathbb{E}$

such that

$$p^{(t)}(u_t\,,\,\cdot\,) \xrightarrow{v} p^{(0)}(u\,,\,\cdot\,) \qquad \textit{in } \mathbb{M}_+(\mathbb{E}') \qquad (t \to \infty)$$

whenever $u_t \to u$ *in* $\mathbb{E}$ *and* $u \in A$. *Suppose also that* $\{\nu^{(t)},\, t \geq 0\}$ *are measures on* $\mathbb{E}$ *such that* $\nu^{(0)}(A^{\mathrm{c}}) = 0$, *and* $\nu^{(t)} \xrightarrow{v} \nu^{(0)}$ *in* $\mathbb{M}_+(\mathbb{E})$. *Then, defining measures* $\mu^{(t)}$ *for* $t \geq 0$ *on* $\mathbb{E} \times \mathbb{E}'$ *as*

$$\mu^{(t)}(du, dx) = \nu^{(t)}(du)p^{(t)}(u\,,\, dx)\,,$$

we have

$$\mu^{(t)} \xrightarrow{v} \mu^{(0)} \qquad \textit{in } \mathbb{M}_+(\mathbb{E} \times \mathbb{E}') \qquad (t \to \infty).$$

Proof. Let $f \in \mathcal{C}_K^+(\mathbb{E} \times \mathbb{E}')$; without loss of generality assume f is supported on $K \times K'$, where $K \in \mathcal{K}(\mathbb{E})$ and $K' \in \mathcal{K}(\mathbb{E}')$. We have

$$\int_{\mathbb{E}\times\mathbb{E}'} \mu^{(t)}(du, dx)\; f(u, x) = \int_{\mathbb{E}} \nu^{(t)}(du) \int_{\mathbb{E}'} p^{(t)}(u\,,\, dx)\; f(u, x).$$

For $t \geq 0$, write

$$\varphi_t(u) = \int_{\mathbb{E}'} p^{(t)}(u\,,\, dx)\; f(u, x)$$

and suppose $u_t \to u_0$ with $u_0 \in A$; we verify that $\varphi_t(u_t) \to \varphi_0(u_0)$. Writing $g_t(x) = f(u_t, x)$, $t \geq 0$, we have $g_t(x_t) \to g_0(x_0)$ whenever $x_t \to x_0 \in \mathbb{E}'$ by the continuity of f. Also, the g_t are uniformly bounded by the bound on f, and $g_t(x) = 0$ for all t whenever $x \notin K'$. Furthermore, without loss of generality we can assume that $p^{(0)}(u_0, \partial K') = 0$. By Lemma A.0.2 (b), we have

$$\varphi_t(u_t) = \int_{\mathbb{E}'} p^{(t)}(u_t\,,\, dx)\; g_t(x) \longrightarrow \int_{\mathbb{E}'} p^{(0)}(u_0\,,\, dx)\; g_0(x) = \varphi_0(u_0).$$

Since the $p^{(t)}$ are substochastic, and $\varphi_t(u) = 0$ for all t whenever $u \notin K$, the φ_t are uniformly bounded by the bound on f. Assume similarly that $\nu(\partial K) = 0$, and recall that $\nu(A^{\mathrm{c}}) = 0$. Apply Lemma A.0.2 (b) once more to conclude as $t \to \infty$

that

$$\int_{\mathbb{E}\times\mathbb{E}'} \mu^{(t)}(du,dx)\ f(u,x) = \int_{\mathbb{E}} \nu^{(t)}(du)\ \varphi_t(u)$$
$$\longrightarrow \int_{\mathbb{E}} \nu^{(0)}(du)\ \varphi_0(u) = \int_{\mathbb{E}\times\mathbb{E}'} \mu^{(0)}(du,x)\ f(u,x). \qquad \square$$

We conclude this section with a result used to verify the existence of the extremal boundary.

Lemma A.0.5. *Suppose P_t, $t \geq 0$ are probability measures on a cs metric space $\mathbb{E}$ such that $P_t \Rightarrow P_0$, and let $A \subset \mathbb{E}$ be measurable. Then there exists a sequence of sets $A_t \downarrow \overline{A}$ such that $P_t(A_t) \to P_0(\overline{A})$.*

Remark. Note that if $P(\partial A) = 0$ then we can take $A_t = \overline{A}$. In the case of distribution functions $F_t \Rightarrow F$ on $\mathbb{R}^m$, taking $A = (-\infty, \boldsymbol{x}]$ and metric $\rho = \rho_\infty$ shows that for any $\boldsymbol{x} \in \mathbb{R}^m$ there exists $\boldsymbol{x}_t \downarrow \boldsymbol{x}$ such that $F_t(\boldsymbol{x}_t) \to F(\boldsymbol{x})$.

Proof. Let ρ be a metric on $\mathbb{E}$, and consider sets $A_\delta = \{x : \rho(x, A) \leq \delta\}$. Recall that $P_0(\partial A_\delta) = 0$ for all but a countable number of choices of δ, since $F(\delta) = P_0(A_\delta) - P_0(\overline{A})$ is a distribution function. First choose $\{\delta_k : k = 1, 2, \ldots\}$ such that $0 < \delta_{k+1} \leq \delta_k \wedge 1/(k+1)$ and $P_0(\partial A_{\delta_k}) = 0$ for all k. Next, let $s_0 = 0$ and take $s_k \geq s_{k-1} + 1$, $k = 1, 2, \ldots$, such that $P_t(A_{\delta_k}) > P_0(\overline{A}) - 1/k$ whenever $t \geq s_k$; this is possible since $P_t(A_{\delta_k}) \to P_0(A_{\delta_k}) \geq P_0(\overline{A})$ for all k. Finally, for $t > 0$ set

$$A(t) = A_{\delta_1}\,\mathbf{1}_{(0,s_1)}(t) + \sum_{k=1}^{\infty} A_{\delta_k}\,\mathbf{1}_{[s_k,s_{k+1})}(t).$$

We claim that $A(t) \downarrow \overline{A}$ and that $P_t(A(t)) \to P_0(\overline{A})$ as $t \to \infty$. It is clear that $A(t) \supset A(t')$ for $t \leq t'$, and $\cap_t A(t) = \cap_k A_{\delta_k} = \overline{A}$. On the one hand, for large t we have $A(t) \subset A_{\delta_k}$ for any k, so

$$\limsup_{t\to\infty}\ P_t(A(t)) \leq \limsup_{t\to\infty}\ P_t(A_{\delta_k}) \leq P_0(A_{\delta_k}).$$

Letting $k \to \infty$ shows that $\limsup_{t\to\infty} P_t(A(t)) \le P_0(\overline{A})$. On the other hand, if $k(t)$ denotes the value of k for which $s_k \le t < s_{k+1}$, then

$$P_t(A(t)) = P_t(A_{\delta_{k(t)}}) > P_0(\overline{A}) - 1/k(t),$$

so $\liminf_t P_t(A(t)) \ge P_0(\overline{A})$. Combining these two inequalities shows that

$$P_t(A(t)) \to P_0(\overline{A}). \qquad \square$$

BIBLIOGRAPHY

[1] S. Asmussen. Extreme value theory for queues via cycle maxima. *Extremes*, 1(2):137–168, 1998.

[2] S. Asmussen. *Applied probability and queues.* Springer Verlag, 2nd edition, 2003.

[3] K.B. Athreya and P. Ney. A new approach to the limit theory of recurrent Markov chains. *Transactions of the American Mathematical Society*, 245:493–501, 1978.

[4] K.B. Athreya and S.G. Pantula. Mixing properties of Harris chains and autoregressive processes. *Journal of applied probability*, 23(4):880–892, 1986.

[5] R.M. Balan and S. Louhichi. Convergence of point processes with weakly dependent points. *Journal of Theoretical Probability*, 22(4):955–982, 2009.

[6] B. Basrak, R.A. Davis, and T. Mikosch. A characterization of multivariate regular variation. *Annals of Applied Probability*, 12(3):908–920, 2002.

[7] B. Basrak, R.A. Davis, and T. Mikosch. Regular variation of GARCH processes. *Stochastic Processes and their Applications*, 99(1):95–115, 2002.

[8] B. Basrak and J. Segers. Regularly varying multivariate time series. *Stochastic Processes and their Applications*, 119(4):1055–1080, 2009.

[9] S.M. Berman. Sojourns and extremes of stationary processes. *The Annals of Probability*, 10(1):1–46, 1982.

[10] P. Bertail, S. Clémençon, and J. Tressou. Extreme values statistics for Markov chains via the (pseudo-) regenerative method. *Extremes*, 12(4):327–360, 2009.

[11] Patrice Bertail and Stphan Clmenon. Regeneration-based statistics for Harris recurrent Markov chains. In Patrice Bertail, Philippe Soulier, and Paul Doukhan, editors, *Dependence in Probability and Statistics*, volume 187 of *Lecture Notes in Statistics*, pages 3–54. Springer New York, 2006.

[12] P. Billingsley. *Convergence of Probability Measures.* Wiley, 1968.

[13] P. Billingsley. *Weak Convergence of Measures: Applications in Probability.* Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1971. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 5.

[14] P. Billingsley. *Probability and Measure.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. J. Wiley & Sons, 1995.

[15] P. Billingsley. *Convergence of Probability Measures.* Wiley-Interscience, 2nd edition, 1999.

[16] N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation.* Encyclopedia of Mathematics and Its Applications. Cambridge University Press, 1989.

[17] M. Borkovec. Extremal behavior of the autoregressive process with ARCH(1) errors. *Stochastic Processes and their Applications*, 85(2):189–207, 2000.

[18] P. Bortot and S. Coles. A sufficiency property arising from the characterization of extremes of Markov chains. *Bernoulli*, 6(1):183–190, 2000.

[19] P. Bortot and S. Coles. Extremes of Markov chains with tail switching potential. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65(4):851–867, 2003.

[20] P. Bortot and J.A. Tawn. Models for the extremes of Markov chains. *Biometrika*, 85(4):851–867, 1998.

[21] M.R. Chernick. A limit theorem for the maximum of autoregressive processes with uniform marginal distributions. *The Annals of Probability*, 9(1):145–149, 1981.

[22] S.G. Coles and J.A. Tawn. Modelling extreme multivariate events. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(2):377–392, 1991.

[23] S.G. Coles, J.A. Tawn, and R.L. Smith. A seasonal Markov model for extremely low temperatures. *Environmetrics*, 5(3):221–239, 1994.

[24] B. Das. *The conditional extreme value model and related topics.* PhD thesis, Cornell University, 2009.

[25] B. Das and S.I. Resnick. Conditioning on an extreme component: Model consistency with regular variation on cones. *Bernoulli*, 17(1):226–252, 2011.

[26] B. Das and S.I. Resnick. Detecting a conditional extreme value model. *Extremes*, 14(1):29–61, 2011.

[27] R. Davis and S. Resnick. Limit theory for moving averages of random variables with regularly varying tail probabilities. *The Annals of Probability*, 13(1):179–195, 1985.

[28] R. Davis and S. Resnick. More limit theory for the sample correlation function of moving averages. *Stochastic Processes and their Applications*, 20(2):257–279, 1985.

[29] R.A. Davis and T. Hsing. Point process and partial sum convergence for weakly dependent random variables with infinite variance. *The Annals of Probability*, 23(2):879–917, 1995.

[30] A.C. Davison and R.L. Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3):393–442, 1990.

[31] Y.A. Davydov. Mixing conditions for Markov chains. *Teoriya Veroyatnostei i ee Primeneniya*, 18(2):321–338, 1973.

[32] L. de Haan and J. de Ronde. Sea and wind: multivariate extremes at work. *Extremes*, 1(1):7–45, 1998.

[33] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer Verlag, 2006.

[34] L. de Haan, S.I. Resnick, H. Rootzén, and C.G. de Vries. Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stochastic Processes and their Applications*, 32(2):213–224, 1989.

[35] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events for insurance and finance*, volume 33. Springer Verlag, 1997.

[36] L. Fawcett and D. Walshaw. Markov chain models for extreme wind speeds. *Environmetrics*, 17(8):795–809, 2006.

[37] P.D. Feigin, M.F. Kratz, and S.I. Resnick. Parameter estimation for moving averages with positive innovations. *The Annals of Applied Probability*, 6(4):1157–1190, 1996.

[38] A.L. Fougères and P. Soulier. Estimation of conditional laws given an extreme component. *Extremes*, 15(1):1–34, 2012.

[39] JL Geluk and L. de Haan. Regular Variation, Extensions and Tauberian Theorems, Volume 40 of CWI Tract. Stichting Mathematisch Centrum, Centrum voor Wiskunde en Informatica, Amsterdam, 1987. Technical report, ISBN 90-6196-324-9, 1987.

[40] P. Glasserman and S.G. Kou. Limits of first passage times to rare sets in regenerative processes. *The Annals of Applied Probability*, 5(2):424–445, 1995.

[41] C.M. Goldie. Implicit renewal theory and tails of solutions of random equations. *The Annals of Applied Probability*, 1(1):126–166, 1991.

[42] L. Haan and S.I. Resnick. Limit theory for multivariate sample extremes. *Probability Theory and Related Fields*, 40(4):317–337, 1977.

[43] G. Haiman, M. Kiki, and M.L. Puri. Extremes of Markov sequences. *Journal of statistical planning and inference*, 45(1-2):185–201, 1995.

[44] J.E. Heffernan and S.I. Resnick. Limit laws for random vectors with an extreme component. *Annals of Applied Probability*, 17(2):537–571, 2007.

[45] J.E. Heffernan and J.A. Tawn. A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 66(3):497–546, 2004.

[46] G. Hooghiemstra and L.E. Meester. Computing the extremal index of special Markov chains and queues. *Stochastic processes and their applications*, 65(2):171–185, 1996.

[47] T. Hsing. On the characterization of certain point processes. *Stochastic processes and their applications*, 26:297–316, 1987.

[48] T. Hsing. Extreme value theory for multivariate stationary sequences. *Journal of Multivariate Analysis*, 29(2):274–291, 1989.

[49] T. Hsing, J. Hüsler, and MR Leadbetter. On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields*, 78(1):97–112, 1988.

[50] H. Hult and F. Lindskog. Extremal behavior of regularly varying stochastic processes. *Stochastic Processes and their applications*, 115(2):249–274, 2005.

[51] D.L. Iglehart. Extreme values in the GI/G/1 queue. *The Annals of Mathematical Statistics*, 43(2):627–635, 1972.

[52] O. Kallenberg. *Foundations of Modern Probability*. Springer Verlag, 1997.

[53] H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131(1):207–248, 1973.

[54] C. Klüppelberg and S. Pergamenchtchikov. Extremal behaviour of models with multivariate random recurrence representation. *Stochastic processes and their applications*, 117(4):432–456, 2007.

[55] M.R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes*. Springer series in statistics. Springer-Verlag, 1983.

[56] MR Leadbetter and H. Rootzén. Extremal theory for stochastic processes. *The Annals of Probability*, 16(2):431–478, 1988.

[57] A.W. Ledford and J.A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187, 1996.

[58] A.W. Ledford and J.A. Tawn. Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(2):475–499, 1997.

[59] A.W. Ledford and J.A. Tawn. Concomitant tail behaviour for extremes. *Advances in Applied Probability*, 30(1):197–215, 1998.

[60] K. Maulik and S. Resnick. Characterizations and examples of hidden regular variation. *Extremes*, 7(1):31–67, 2004.

[61] K. Maulik, S. Resnick, and H. Rootzén. Asymptotic independence and a network traffic model. *Journal of Applied Probability*, 39(4):671–699, 2002.

[62] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Communications and control engineering. Cambridge University Press, 2009.

[63] T. Mikosch. How to model multivariate extremes if one must? *Statistica Neerlandica*, 59(3):324–338, 2005.

[64] T. Mikosch and S. Resnick. Activity rates with very heavy tails. *Stochastic processes and their applications*, 116(2):131–155, 2006.

[65] T. Mikosch and C. Starica. Limit theory for the sample autocorrelations and extremes of a GARCH$(1,1)$ process. *Annals of Statistics*, 28(5):1427–1451, 2000.

[66] J. Neveu. Processus ponctuels. In *Ecole dEté de Probabilités de Saint-Flour VI-1976*, volume 598, pages 249–445. Springer Berlin/Heidelberg, 1977.

[67] SY Novak. Multilevel clustering of extremes. *Stochastic processes and their applications*, 97(1):59–75, 2002.

[68] G.L. O'Brien. Extreme values for stationary and Markov sequences. *The Annals of Probability*, 15(1):281–291, 1987.

[69] R. Perfekt. Extremal behaviour of stationary Markov chains with applications. *The Annals of Applied Probability*, 4(2):529–548, 1994.

[70] R. Perfekt. Extreme value theory for a class of Markov chains with values in $\mathbb{R}^d$. *Advances in Applied Probability*, 29(1):138–164, 1997.

[71] J. Pickands. Statistical inference using extreme order statistics. *the Annals of Statistics*, 3:119–131, 1975.

[72] J. Pickands. Multivariate extreme value distributions. In *Proceedings 43rd Session International Statistical Institute*, pages 859–878, 1981.

[73] S. Resnick. Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5(4):303–336, 2002.

[74] S. Resnick and H. Rootzén. Self-similar communication models and very heavy tails. *Annals of Applied Probability*, 10(3):753–778, 2000.

[75] S.I. Resnick. Point processes, regular variation and weak convergence. *Advances in Applied Probability*, 18(1):66–138, 1986.

[76] S.I. Resnick. *Extreme Values, Regular Variation, and Point Processes.* Springer Verlag, 2007.

[77] S.I. Resnick. *Heavy-tail Phenomena: Probabilistic and Statistical Modeling.* Springer Verlag, 2007.

[78] S.I. Resnick. Multivariate regular variation on cones: application to extreme values, hidden regular variation and conditioned limit laws. *Stochastics: An International Journal of Probability and Stochastics Processes*, 80(2-3):269–298, 2008.

[79] S.I. Resnick and D. Zeber. Asymptotics of Markov kernels and the tail chain. *Advances in Applied Probability*, 45(1), 2013. To appear.

[80] C.Y. Robert. Inference for the limiting cluster size distribution of extreme values. *The Annals of Statistics*, 37(1):271–310, 2009.

[81] H. Rootzén. Maxima and exceedances of stationary Markov chains. *ADV. APPL. PROB.*, 20(2):371–390, 1988.

[82] J. Segers. Functionals of clusters of extremes. *Advances in Applied Probability*, 35(4):1028–1045, 2003.

[83] J. Segers. Approximate distributions of clusters of extremes. *Statistics and Probability Letters*, 74(4):330–336, 2005.

[84] J. Segers. Multivariate regular variation of heavy-tailed Markov chains. *Arxiv preprint math/0701411*, 2007.

[85] R. Serfozo. High-level exceedances of regenerative and semi-stationary processes. *Journal of Applied Probability*, 17(2):423–431, 1980.

[86] S. Sisson and S. Coles. Modelling dependence uncertainty in the extremes of Markov chains. *Extremes*, 6(4):283–300, 2003.

[87] R.L. Smith. The extremal index for a Markov chain. *Journal of applied probability*, 29(1):37–45, 1992.

[88] R.L. Smith, J.A. Tawn, and S.G. Coles. Markov chain models for threshold exceedances. *Biometrika*, 84(2):249, 1997.

[89] S. Yun. The extremal index of a higher-order stationary Markov chain. *Annals of Applied Probability*, 8(2):408–437, 1998.

[90] S. Yun. The distributions of cluster functionals of extreme events in a dth-order Markov chain. *Journal of Applied Probability*, 37(1):29–44, 2000.